

BRUNO EDUARDO DONNI

**ANÁLISE DOS PADRÕES DE USO DO TRANSPORTE PÚBLICO INDIVIDUAL EM
GRANDE CENTRO URBANO E DOS IMPACTOS CAUSADOS PELA ENTRADA
DE UM CONCORRENTE PRIVADO NESTE MERCADO**

São Paulo

2019

BRUNO EDUARDO DONNI

**ANÁLISE DOS PADRÕES DE USO DO TRANSPORTE PÚBLICO INDIVIDUAL EM
GRANDE CENTRO URBANO E DOS IMPACTOS CAUSADOS PELA ENTRADA
DE UM CONCORRENTE PRIVADO NESTE MERCADO**

Trabalho de Formatura apresentado à Escola
Politécnica da Universidade de São Paulo para
a obtenção do diploma de Engenheiro de
Produção

São Paulo

2019

BRUNO EDUARDO DONNI

**ANÁLISE DOS PADRÕES DE USO DO TRANSPORTE PÚBLICO INDIVIDUAL EM
GRANDE CENTRO URBANO E DOS IMPACTOS CAUSADOS PELA ENTRADA
DE UM CONCORRENTE PRIVADO NESTE MERCADO**

Trabalho de Formatura apresentado à Escola
Politécnica da Universidade de São Paulo para
a obtenção do diploma de Engenheiro de
Produção

Orientador: Prof. Dr. Daniel de Oliveira Mota

São Paulo

2019

Catálogo-na-publicação

Donni, Bruno Eduardo

Análise dos padrões de uso do transporte público individual em grande centro urbano e dos impactos causados pela entrada de um concorrente privado neste mercado / B. E. Donni -- São Paulo, 2019.
95 p.

Trabalho de Formatura - Escola Politécnica da Universidade de São Paulo. Departamento de Engenharia de Produção.

1. Transporte público individual 2. Transporte privado individual 3. Uber 4. Táxi 5. Mobilidade urbana I. Universidade de São Paulo. Escola Politécnica. Departamento de Engenharia de Produção II. t.

À minha família e amigos.

AGRADECIMENTOS

Ao Professor Doutor Daniel de Oliveira Mota, por todas as longas discussões, paciência, e ensinamentos transmitidos ao longo da orientação deste projeto.

A todos os professores do Departamento de Engenharia de Produção da Escola, por me tornarem apto a formar Engenheiro de Produção.

À minha família, pelo apoio incondicional durante a realização deste trabalho e em todas as etapas da minha vida.

Aos amigos que estiveram ao meu lado durante essa jornada.

“To strive, to seek, to find, and not to yield”

Alfred, Lord Tennyson

RESUMO

O mercado de transporte público de passageiros está em transformação devido ao surgimento de concorrentes privados. Este trabalho buscou analisar o impacto que a ascensão destas empresas teve nos padrões de uso de corridas deste serviço. Para isso, foram utilizados dados de todas as corridas de táxi realizadas em Nova Iorque entre 2009 e 2018 e de Uber entre 2015 e 2018, divulgadas pela política de dados abertos da cidade. Dado ao grande volume de informações, as análises foram feitas com ferramentas modernas e especializadas (Alteryx e Tableau). Este trabalho identificou que, embora tenha ocorrido uma acentuada queda no número de corridas de táxi, não há diferenças significativas nos padrões de deslocamento deste tipo de transporte em diferentes dimensões, sendo elas: (1) a distância e a duração das viagens; (2) a sazonalidade; (3) a distribuição espacial; e (4) as relações de origens e destino. Adicionalmente, a entrada da Uber trouxe diferentes consequências dependendo das zonas de atuação analisadas, com (1) aumento da concorrência em áreas que já possuíam presença forte dos táxis e (2) crescimento do total de corridas em localidades com presença fraca.

Palavras-chave: transporte público individual, transporte público privado, táxi, Uber, mobilidade urbana, mobilidade humana.

ABSTRACT

The ride sharing market is in a transformation process due to the emergence of for-hire vehicles private companies. This paper analyses the impact of that rise on the usage patterns of taxi cabs services. It examines data from every Yellow Taxi trip that happened in New York City from 2009 to 2018 and every Uber trip from 2015 to 2018, made public by the city's open data policy. Due to the large amount of information, the analyses were made with modern and specialized tools (Alteryx and Tableau). This study has identified that, although there has been a significant drop in the number of taxi rides, there are no significant differences in the patterns of those trips in different dimensions, namely: (1) the distance and duration of rides; (2) seasonality; (3) spatial distribution; and (4) the origin and destination relationships. Moreover, Uber's entry had different consequences depending on the analyzed zones, with a (1) rise in competition in areas with previous strong Yellow Taxi presence; and (2) a growth in the total number of trips in locations with weak presence.

Keywords: for-hire vehicles, ride sharing, taxi, Uber, urban mobility, human mobility.

LISTA DE FIGURAS

Figura 01: Recorte da página utilizada para <i>download</i> dos dados, mostrando informações de 2018	30
Figura 02: Mapa da cidade de Nova Iorque segmentado pelas zonas de táxi da cidade. As cores representam os <i>boroughs</i> nos quais essas regiões estão inseridas.....	32
Figura 03: Framework com possíveis áreas de aplicação de <i>big data</i> no ciclo de vida de um produto.....	35
Figura 04: Parâmetros da arquitetura do sistema TaxiVis.....	41
Figura 05: Visualização para o usuário da seleção de parâmetros analisados e resultados do TaxiVis	42
Figura 06: Número de corridas de <i>Yellow Taxi</i> realizadas na cidade em 2011 e 2012	43
Figura 07: Número de <i>pick-ups</i> e <i>drop-offs</i> realizados ao longo da primeira semana de maio de 2011 para diferentes bairros do <i>borough</i> de Manhattan	43
Figura 08: Distribuição espacial e por horário da quantidade de corridas realizadas em Manhattan em diferentes segundas feiras para os anos analisados	44
Figura 09: Critérios de pesquisa do presente trabalho.....	47
Figura 10: Exemplo de fluxo do Alteryx, mostrando o tratamento e a junção de três bases de dados.....	54
Figura 11: Tela do Tableau com visualização geográfica de dados da cidade de Nova Iorque	56
Figura 12: quadro ilustrativo com o processo de preparação dos dados	58
Figura 13: Fluxo do Alteryx que representa o tratamento inicial dos dados de táxi para o ano de 2009	59
Figura 14: Fluxo do Alteryx que representa o tratamento inicial dos dados de <i>FHV</i> para o ano de 2018	60
Figura 15: Fluxo do Alteryx que representa o tratamento adicional realizado nos dados de táxi para incluir as zonas de <i>pick-up</i> para o ano de 2009	61
Figura 16: Número de corridas mês a mês realizadas em Nova Iorque por categoria de transporte individual.....	62
Figura 17: Número de corridas ano a ano realizadas em Nova Iorque por categoria de transporte individual.....	63
Figura 18: Comparação entre crescimento do número de corridas de Uber e queda das corridas de táxi em Nova Iorque entre 2015 e 2018.....	63

Figura 19: Fluxo do Alteryx utilizado para calcular a distância e a duração das viagens de táxi.	66
Figura 20: Fluxo do Alteryx utilizado para calcular a distância e a duração das viagens de Uber.	67
Figura 21: Fluxo do Alteryx utilizado para calcular a sazonalidade por dia da semana e mês das viagens de táxi.	68
Figura 22: Fluxo do Alteryx utilizado para calcular a sazonalidade por dia da semana e mês das viagens de Uber.	69
Figura 23: Fluxo do Alteryx utilizado para visualizar as viagens por zona, dia da semana e horário das viagens de táxi e Uber	70
Figura 24: Conexão da base gerada com as viagens por zona, dia da semana e horário das viagens de táxi e Uber com a base de zona de táxi definidas pela <i>TLC</i> .	71
Figura 25: Fluxo do Alteryx utilizado para analisar a recorrência de origem-destino das viagens de táxi e Uber.	72
Figura 26: Fluxo do Alteryx utilizado para analisar a evolução das corridas de táxi que começaram ou terminaram em Manhattan	73
Figura 27: Fluxo do Alteryx utilizado para analisar a evolução das corridas de táxi que começaram e terminaram em Manhattan.	74
Figura 28: Fluxo do Alteryx utilizado para analisar a evolução das corridas de Uber que começaram e/ou terminaram em Manhattan e as corridas que começaram e terminaram neste <i>borough</i> .	74
Figura 29: Mediana das distâncias percorridas pelos táxis mês a mês.	75
Figura 30: Mediana das durações das corridas de táxi e Uber mês a mês.	76
Figura 31: Corridas de Uber por dia da semana e por ano.	77
Figura 32: Corridas de táxi por dia da semana e por ano.	77
Figura 33: Corridas de Uber por mês e por ano.	78
Figura 34: Corridas de táxi por mês e por ano	79
Figura 35: Indicação do botão utilizado para visualizar os mapas em tela cheia dentro da página do Tableau Public com a análise do número de <i>pick-ups</i> de táxi por dia da semana	80
Figura 36: Indicação do botão utilizado para selecionar o dia da semana desejado na visualização dentro da página do Tableau Public com a análise do número de <i>pick-ups</i> de táxi por dia da semana.	80
Figura 37: Diagrama de Sankey para as corridas de táxi em 2009, representando a origem e o destino das viagens entre os <i>boroughs</i> de Nova Iorque.	82

Figura 38: Diagrama de Sankey para as corridas de táxi em 2013, representando a origem e o destino das viagens entre os <i>boroughs</i> de Nova Iorque.....	83
Figura 39: Diagrama de Sankey para as corridas de táxi em 2018, representando a origem e o destino das viagens entre os <i>boroughs</i> de Nova Iorque.....	84
Figura 40: Diagrama de Sankey para as corridas de Uber em 2018, representando a origem e o destino das viagens entre os <i>boroughs</i> de Nova Iorque.....	85
Figura 41: Evolução das corridas que começam e/ou terminam em Manhattan por tipo de transporte individual.....	87
Figura 42: Evolução das corridas que começam e terminam em Manhattan por tipo de transporte individual.....	87

LISTA DE TABELAS

Tabela 01: Quadro-resumo com as informações divulgadas ano a ano para táxis que foram analisadas no estudo	50
Tabela 02: Quadro-resumo com as informações divulgadas ano a ano para <i>FHVs</i> que foram analisadas no estudo	51
Tabela 03: <i>Links</i> para visualização da análise do número de viagens por zona de táxi, dia da semana e horário	79

LISTA DE SIGLAS E ABREVIACES

NYSE	<i>New York Stock Exchange</i>
EUA	Estados Unidos da Amrica
TLC	<i>Taxi and Limousine Commission</i>
FHV	<i>For-Hire Vehicles</i>
EWR	Aeroporto Internacional de Newark Liberty
BOL	Comeo da vida (<i>beginning-of-life</i>)
MOL	Meio da vida (<i>middle-of-life</i>)
EOL	Fim da vida (<i>end-of-life</i>)

SUMÁRIO

1. INTRODUÇÃO	27
1.1. Problema estudado.....	28
1.2. Objetivo do trabalho	28
1.3. Justificativa.....	29
2. REVISÃO BIBLIOGRÁFICA	33
2.1. <i>Big data</i> e suas áreas de aplicação.....	33
2.2. Transporte individual de passageiros em grandes centros urbanos.....	38
3. METODOLOGIA.....	47
3.1. Metodologia de pesquisa	47
3.1.1. Apresentação da metodologia escolhida.....	47
3.2. Metodologia do projeto	48
3.2.1. Fonte de dados	48
3.2.1.1. Descrição das variáveis de estudo	49
3.2.1.2. Restrições dos dados.....	51
3.2.2. Ferramentas de análise utilizadas	52
3.2.2.1. Alteryx Designer	53
3.2.2.2. Tableau Desktop	55
4. PROJETO	57
4.1. Preparação das bases de dados	57
4.1.1. Dados de táxi	58
4.1.2. Dados de <i>FHV</i>	59
4.1.3. Dados de táxi com <i>pick-up</i> e <i>drop-off</i> por zona de táxi.....	60
4.2. Análise descritiva dos dados.....	61
4.3. Descrição do processo de análise	64
4.3.1. Distância e duração das viagens	64

4.3.2. Sazonalidade por dia da semana e mês	67
4.3.3. Número de viagens por zona de táxi, dia da semana e horário	69
4.3.4. Localização de <i>pick-up</i> e <i>drop-off</i>	71
4.3.5. Evolução das corridas em Manhattan	72
4.4. Resultado das análises.....	75
4.4.1. Distância e duração das viagens.....	75
4.4.2. Sazonalidade por dia da semana e mês	76
4.4.2.1. Sazonalidade por dia da semana	76
4.4.2.1. Sazonalidade por mês.....	78
4.4.3. Número de viagens por zona de táxi, dia da semana e horário	79
4.4.4. Localização de <i>pick-up</i> e <i>drop-off</i>	81
4.4.5. Evolução das corridas em Manhattan	86
4.4.5.1. Evolução das corridas que começam e/ou terminam em Manhattan	86
4.4.5.1. Evolução das corridas que começam e terminam em Manhattan	87
5. CONCLUSÃO	89
5.1. Trabalhos futuros	90
REFERÊNCIAS BIBLIOGRÁFICAS	91

1. INTRODUÇÃO

O transporte público individual é uma indústria que está sendo transformada pela ascensão de empresas privadas. Estas trouxeram uma concorrência sem precedentes para o mercado (Cramer e Krueger, 2016).

O transporte público individual é o serviço remunerado de transporte de passageiros realizado por meio de viagens individualizadas ou em pequenos grupos e abertas ao público. Tradicionalmente, esse termo refere-se ao transporte de táxi, que é um meio de locomoção autorizado pelos governos sob o regime de concessões de licenças (ou alvarás) aos taxistas (Gallick e Sisk, 1987).

Essa forma de transporte é um meio antigo de locomoção utilizado em grandes centros urbanos. Segundo Dempsey (1996), grandes cidades americanas iniciaram o movimento de regulamentação dos táxis nos anos de 1920, indicando que desde aquela época o taxi é um meio de transporte relevante nos ambientes urbanos. Na maioria das cidades, o número de licenças de táxi é controlado pelo governo, de modo que somente um número limitado de pessoas pode realizar este serviço (Angrist et al., 2017).

Nos últimos 10 anos, observou-se a ascensão de empresas privadas que exploram esse mesmo nicho. Essas empresas realizam o transporte de passageiros por meio de uma rede credenciada de motoristas, que podem possuir carro próprio ou alugado, sem a necessidade de licença do governo. O usuário a ser transportado, por meio de um aplicativo, seleciona a origem e o destino desejados para a corrida, que é transmitida para motoristas que aceitam ou não realizar o transporte. A empresa conecta os dois interessados e transaciona o pagamento entre os usuários do sistema, recebendo em troca uma comissão sobre o valor pago pela corrida.

A maior empresa desse ramo é a Uber (Wallsten, 2015). A Uber Technologies Inc. foi fundada em 2009, possui atuação em mais de 700 cidades, 57 países em todos os continentes, e até o final de 2018 já realizou um total de mais de 10 bilhões de corridas (Uber, 2019). Em maio de 2019, a empresa abriu o seu capital nos Estados Unidos, na NYSE (*New York Stock Exchange*), sendo avaliada em 82,4 bilhões de dólares (New York Times, 2019).

Como foi observado por Cannon e Summers (2014), essas empresas trazem benefícios significativos para a economia, para o meio ambiente e para os empreendedores, incluindo aumento no número de empregos (por meio da criação de novos postos de trabalho). Entretanto, governos de municípios onde estas empresas entram, em geral, reagem com resistência, criando barreiras jurídicas para a atuação dos novos entrantes, chegando, em alguns casos, a proibir a atuação dos mesmos. Para Wyman (2017), não há razões econômicas que justifiquem

diferenças na legislação entre o transporte público individual e o privado, dada a mesma natureza do serviço prestado. Dessa forma, a atuação legislativa dos agentes públicos, ao impor restrições a empresas como a Uber, não está cumprindo a sua função de compensar externalidades negativas e maximizar o bem-estar da sociedade, mas sim exercendo um papel político protecionista.

Além disso, a competição com a Uber pode mudar características do serviço de táxi. Wallsten (2015) argumenta que a crescimento da Uber está associado à diminuição das reclamações sobre o serviço oferecido pelos taxistas aos passageiros, incluindo máquinas de cartão de crédito quebradas, falta de ar condicionado e aquecimento, além de tratamento inadequado e uso de celular ao volante.

Ainda segundo o mesmo autor, uma iniciativa desta empresa para evitar prejuízo por questões regulatórias seria a divulgação de estudos sobre os efeitos positivos que a presença da companhia nos ambientes aos quais a empresa está inserida. Entretanto, não foi encontrado na literatura estudos sobre os impactos da entrada deste competidor nas viagens dos usuários de transporte público individual em grandes centros urbanos.

1.1. Problema estudado

Estima-se que a ascensão do transporte privado individual interfere, de alguma forma, em sua versão pública, dado o aumento de competição pelos mesmos passageiros (no caso do mercado já existente), ou de novos passageiros (no caso de ampliação deste mercado). Ao aumentar o número de carros oferecendo esse serviço de transporte em um grande centro urbano, pode-se acreditar na alteração da dinâmica da mobilidade urbana no local, como mudanças na distância e duração das viagens, na sazonalidade dos trajetos, e nos caminhos cursados. Segundo Liu et al. (2012), os padrões das trajetórias percorridas pelos táxis em uma cidade refletem as características da mobilidade humana da mesma.

Para se aprofundar no entendimento desta interferência ocorrida pela entrada de agente privado, o presente estudo busca analisar os impactos nos tipos de uso de veículos de transporte público individual licenciados pelo serviço público (taxi) antes e depois da ascensão de concorrentes privados (Uber) em grandes centros urbanos.

1.2. Objetivo do trabalho

Com o objetivo de investigar o problema apresentado, este trabalho buscou identificar diferenças no padrão de uso de corridas de *Yellow Taxis* (concessão de taxi da prefeitura de Nova Iorque – EUA) antes e depois da entrada da Uber na cidade de Nova Iorque para usuários

de transporte público individual. O atendimento deste objetivo favoreceu, também, a identificação de fatores secundários relacionados a mobilidade dos usuários na cidade, que foram evidenciados ao longo do trabalho.

1.3. Justificativa

O estudo da mobilidade urbana e humana é fundamental para aspectos importantes dos grandes centros urbanos, como: controle de epidemias, planejamento urbano, sistemas de previsão de tráfego e aplicações de rede (Zhao et al., 2015).

Entende-se que a entrada de um concorrente no mercado de transporte público individual, ao evitar um monopólio dos taxistas, pode acarretar modificações nos padrões de uso dos táxis, o que, por sua vez, pode influenciar na dinâmica da mobilidade urbana e humana de uma cidade.

A compreensão de como o fenômeno do surgimento de novas modalidades de transporte altera os padrões de mobilidade em grandes centros urbanos é benéfico para as diversas esferas da sociedade, como pode ser observado, de maneira não limitada, nos seguintes exemplos: (1) auxílio aos agentes regulatórios que devem maximizar o bem-estar social ao atualizar as políticas públicas para o surgimento de novos agentes nos meios de transporte de grandes centros urbanos; (2) servir de referência para a atualização de estudos sobre mobilidade urbana antes do surgimento dessas novas formas de transporte de passageiros; (3) auxiliar no desenho dos meios de transporte em estudos sobre cidades inteligentes ou outras formas de planejamento urbano.

Além disso, o trabalho é aderente com a formação de engenharia de produção, pertencendo a duas distintas áreas de profissão, como definidas pela Associação Brasileira de Engenharia de Produção (ABEPRO, 2019). Essas áreas são a de Projeto e Análise de Sistemas Logísticos (2.3.), por avaliar o transporte e os deslocamentos de pessoas em Nova Iorque; e a Análise de Demanda (3.6.), por analisar a procura por transporte individual na cidade.

O direcionamento do escopo do trabalho para a cidade de Nova Iorque e a consequente comparação entre os *Yellow Taxis* e Ubers, foi feita em virtude da disponibilidade de dados públicos que serão utilizados nas análises.

Em 2012, a câmara municipal da cidade aprovou uma lei para tornar os dados do governo abertos ao público, em uma tentativa de tornar a administração local mais transparente e de encorajar o público a identificar maneiras de melhorar a comunidade local (Okamoto, 2016). Como parte dessa iniciativa, a *Taxi and Limousine Commission (TLC)*, divulga dados com o registro de todas as corridas de transporte público individual e privado que acontecem

na cidade por meio de seu portal na internet, com bases de dados fornecidas por mês, como visto na figura 01.

TLC Trip Record Data

Expand All Collapse All

The yellow and green taxi trip records include fields capturing pick-up and drop-off dates/times, pick-up and drop-off locations, trip distances, itemized fares, rate types, payment types, and driver-reported passenger counts. The data used in the attached datasets were collected and provided to the NYC Taxi and Limousine Commission (TLC) by technology providers authorized under the Taxicab & Livery Passenger Enhancement Programs (TPEP/LPEP). The trip data was not created by the TLC, and TLC makes no representations as to the accuracy of these data.

The For-Hire Vehicle ("FHV") trip records include fields capturing the dispatching base license number and the pick-up date, time, and taxi zone location ID (shape file below). These records are generated from the FHV Trip Record submissions made by bases. Note: The TLC publishes base trip record data as submitted by the bases, and we cannot guarantee or confirm their accuracy or completeness. Therefore, this may not represent the total amount of trips dispatched by all TLC-licensed bases. The TLC performs routine reviews of the records and takes enforcement actions when necessary to ensure, to the extent possible, complete and accurate information.

2018

January - Yellow Taxi Trip Records (CSV) - Green Taxi Trip Records (CSV) - FHV Trip Records (CSV) February - Yellow Taxi Trip Records (CSV) - Green Taxi Trip Records (CSV) - FHV Trip Records (CSV) March - Yellow Taxi Trip Records (CSV) - Green Taxi Trip Records (CSV) - FHV Trip Records (CSV) April - Yellow Taxi Trip Records (CSV) - Green Taxi Trip Records (CSV) - FHV Trip Records (CSV) May - Yellow Taxi Trip Records (CSV) - Green Taxi Trip Records (CSV) - FHV Trip Records (CSV) June - Yellow Taxi Trip Records (CSV) - Green Taxi Trip Records (CSV) - FHV Trip Records (CSV)	July - Yellow Taxi Trip Records (CSV) - Green Taxi Trip Records (CSV) - FHV Trip Records (CSV) August - Yellow Taxi Trip Records (CSV) - Green Taxi Trip Records (CSV) - FHV Trip Records (CSV) September - Yellow Taxi Trip Records (CSV) - Green Taxi Trip Records (CSV) - FHV Trip Records (CSV) October - Yellow Taxi Trip Records (CSV) - Green Taxi Trip Records (CSV) - FHV Trip Records (CSV) November - Yellow Taxi Trip Records (CSV) - Green Taxi Trip Records (CSV) - FHV Trip Records (CSV) December - Yellow Taxi Trip Records (CSV) - Green Taxi Trip Records (CSV) - FHV Trip Records (CSV)
---	---

Figura 01: Recorte da página utilizada para *download* dos dados, mostrando informações de 2018.

Fonte: website da TLC.

O ambiente competitivo de transporte público privado na cidade é formado por 3 principais agentes: os (1) *Yellow Taxis*, os (2) *For-Hire Vehicles (FHVs)* e os (3) *Green Taxis*.

O (1) *Yellow Taxi*, ou táxi amarelo, corresponde ao tradicional serviço de táxi ou transporte público individual da cidade. Os (2) *FHVs* compõe a categoria que engloba o

transporte privado individual na cidade, sendo a empresa Uber a sua provedora de maior relevância, embora outras também estejam nesta categoria, como a Lyft, a Via e a Juno. A empresa iniciou suas operações em Nova Iorque no mês de maio de 2011 (Poulsen et al., 2016). Segundo dados apresentados pela própria empresa, essa entrada de mercado contou com apenas 100 carros e foi a primeira expansão para fora de São Francisco, onde foi originada (Uber, 2011). Já o (3) *Green Taxi*, ou táxi verde, é uma categoria criada pela *TLC* em agosto de 2013 para buscar passageiros em regiões pouco atendidas pelos táxis regulares (Poulsen et al., 2016). Entretanto, os táxis verdes não conseguiram se tornar relevantes na cidade, pois iniciaram as suas operações concomitantemente ao fortalecimento dos *FHVs*. Segundo Bhairavi Desai, o diretor executivo do Sindicato dos Taxistas de Nova Iorque, a Uber e a Lyft são responsáveis pela quantidade pouco representativa de táxis verdes devido à forte concorrência pelo mesmo mercado de corridas, que acabou sendo dominado por estas empresas (New York Times, 2018).

As informações divulgadas pela agência são compostas por todas as corridas de táxi amarelo desde 2009, as de táxi verde desde o lançamento (agosto de 2013) e as de *FHVs* a partir do início de 2015. Os dados de *Green Taxi* foram despriorizados no presente estudo em função de sua baixa relevância no ambiente competitivo da cidade.

A cidade é um grande centro urbano estadunidense, com cerca de 8,4 milhões de habitantes em 2018 (New York City Department of City Planning, 2018). Ela é dividida em 5 *boroughs*, equivalentes às grandes regiões da cidade. Além disso, essas regiões são divididas em 263 zonas de táxi diferentes, que são grupos de endereços definidos pela *TLC* para demarcar as zonas de *pick-up* e *drop-off* das corridas de táxi e *FHV* nas bases fornecidas pela mesma. As zonas de táxi são baseadas nos bairros da cidade.

Pode-se observar na figura 02 a divisão de zonas de táxi agrupadas por *borough* em Nova Iorque, sendo eles Bronx, Brooklyn, Manhattan, Queens e Staten Island. Há uma denominação específica para o Aeroporto Internacional de Newark Liberty (EWR) pois ele não está contido nos *boroughs*.



Figura 02: Mapa da cidade de Nova Iorque segmentado pelas zonas de táxi da cidade. As cores representam os *boroughs* nos quais essas regiões estão inseridas.
Fonte: zonas de táxi da TLC, mapa elaborado pelo autor.

1.4. Estrutura

O trabalho foi dividido em 5 capítulos: (1) introdução, (2) revisão bibliográfica, (3) metodologia, (4) projeto e (5) conclusão.

A introdução apresenta a temática do transporte público individual e privado e delimita o escopo do trabalho – os padrões de uso de *Yellow Taxis* e *Ubers* em Nova Iorque. Na revisão bibliográfica são abordados dois temas principais: Big Data, ressaltando as diferentes áreas nas quais ele pode ser empregado; e o transporte de táxi em grandes centros urbanos, que evidencia outros estudos realizados com informações similares às deste documento. No capítulo de metodologia, descreve-se as técnicas de pesquisa empregadas e o método de projeto que será utilizado. Em seguida, no capítulo de projeto, aplica-se os métodos apresentados no capítulo anterior para realizar as análises dos dados e mostrar os resultados obtidos. E, por fim, na conclusão, analisa-se o trabalho e comenta-se os resultados e as limitações encontrados.

2. REVISÃO BIBLIOGRÁFICA

Para ampla (porém não exaustiva) compreensão do estudo dos impactos nos padrões de uso causados pela inclusão do transporte privado individual quando comparado ao transporte público foram levantadas algumas das principais referências empregadas por autores que investigaram temas relacionados ao assunto. Estas referências foram organizadas em duas grandes áreas: (1) *big data*; onde foram apresentadas técnicas aplicadas em um elevado volume de dados com informações tanto espaciais quanto temporais comuns a este tipo de análise; e (2) transporte de táxi; dentro do contexto de grandes centros urbanos (megacidades), uma vez que tem-se apoiado iniciativas cada vez mais fortes de se substituir o transporte em veículos próprios por alternativas de caráter coletivo, na qual o uso de táxis vem sendo amplamente incentivado, atraindo atenção do mercado.

2.1. *Big data* e suas áreas de aplicação

Segundo Gandomi e Haider (2015), *big data* é um termo que apresenta definição imprecisa, pois está em constante evolução. De maneira geral, ela é definida por meio de uma série de características relacionadas aos dados que estão nela inseridos. A primeira definição utilizada em larga escala, proposta por Laney (2001) descreve *big data* como dados que apresentam características específicas resumidas no que são popularmente propostas por “3 Vs”. Eles são: (1) Volume, referente à quantidade física destinada para uma eventual necessidade de armazenagem destes dados e, conseqüentemente relativo ao espaço que estes dados venham a ocupar; (2) Variedade, indicativa da estrutura diversa de um conjunto de dados; e finalmente, (3) Velocidade, que se refere a elevada taxa com que esses dados são gerados e armazenados dentro de um ambiente. Entretanto, para Gandomi e Haider (2015), pode-se recorrer a uma definição atualizada, ao incluir-se outras dimensões para os dados e conseqüentemente, de maneira popular, completando os “6 Vs”. Essas dimensões são: (4) Veracidade, relativa a confiabilidade dos dados que estão sendo utilizados e a necessidade frequente de lidar com dados com pouca qualidade, como informações extraídas de redes sociais; (5) Variabilidade (ou complexidade), dadas as possíveis variações na velocidade das taxas nas quais esses dados são gerados; e por fim, (6) Valor, devido à baixa densidade de importância nos dados individuais e a necessidade de realizar análises utilizando o grande volume das informações disponíveis.

Usando-se dos conceitos apresentados anteriormente, Liu et al. (2012), afirma que o extensivo uso de grandes volumes de dados e a explosão de *big data* estão bem documentados por meio de estudos e têm facilitado a realização de análises complexas devido a

disponibilidade de informações em maiores escala e nível de detalhe. Entretanto, o emprego de técnicas relacionadas a *big data* está sujeito a uma série de erros, prejudicando as conclusões alcançadas. Isso pode ocorrer devido ao uso inadequado das técnicas (qualidade das análises) ou à baixa qualidade dos dados (seja pelas formas de coleta, armazenamento ou tratamentos primários). Desta forma, recomenda-se que todo tratamento de dados deve ser realizado com o maior rigor formal possível. Liu et al. (2012) ainda afirma que, dentre as fontes de erro no tratamento dos dados, destacam-se: (1) a coleta não autêntica dos dados, (2) a não completude ou baixa representatividade das informações, (3) o ruído dos dados, (4) a falta de confiança nas informações disponíveis e até mesmo (5) questões éticas. Por conta disso, deve-se estar atento a origem das informações e a melhor forma de tratar os dados antes de divulgar grandes afirmações a partir das análises.

Dito isso, ao utilizar-se os dados de maneira correta, as análises de *big data* permitem fazer estudos em diversas e abrangentes áreas do conhecimento, como pode ser visto nos seguintes exemplos.

Para McAfee e Brynjolfsson (2012), explorar os dados gerados pelos diversos fluxos de informação dentro de uma empresa pode melhorar radicalmente a performance de uma companhia. Isto ocorre, pois, ao usar dados em grande quantidade, rápida produção e altíssima variedade, gestores podem tomar medidas mais rápidas e precisas de como gerir os seus negócios. Ainda segundo McAfee e Brynjolfsson (2012), companhias com processos de decisão embasados em dados chegam a ser 5% mais produtivas e 6% mais lucrativas que os seus competidores. Eles argumentam que, para as empresas conseguirem utilizar a *big data* de maneira eficaz, elas precisam passar por 5 diferentes desafios gerenciais: (1) liderança com metas e objetivos bem definidos; (2) gerenciamento dos talentos responsáveis por gerar as informações e os *insights* por meio dos dados; (3) ter ferramentas tecnológicas adequadas; (4) criar um processo de decisão que utilize as informações e os *insights*, e não apenas o *expertise* de sua liderança; e, por fim, (5) criar uma cultura “*data-driven*” na companhia.

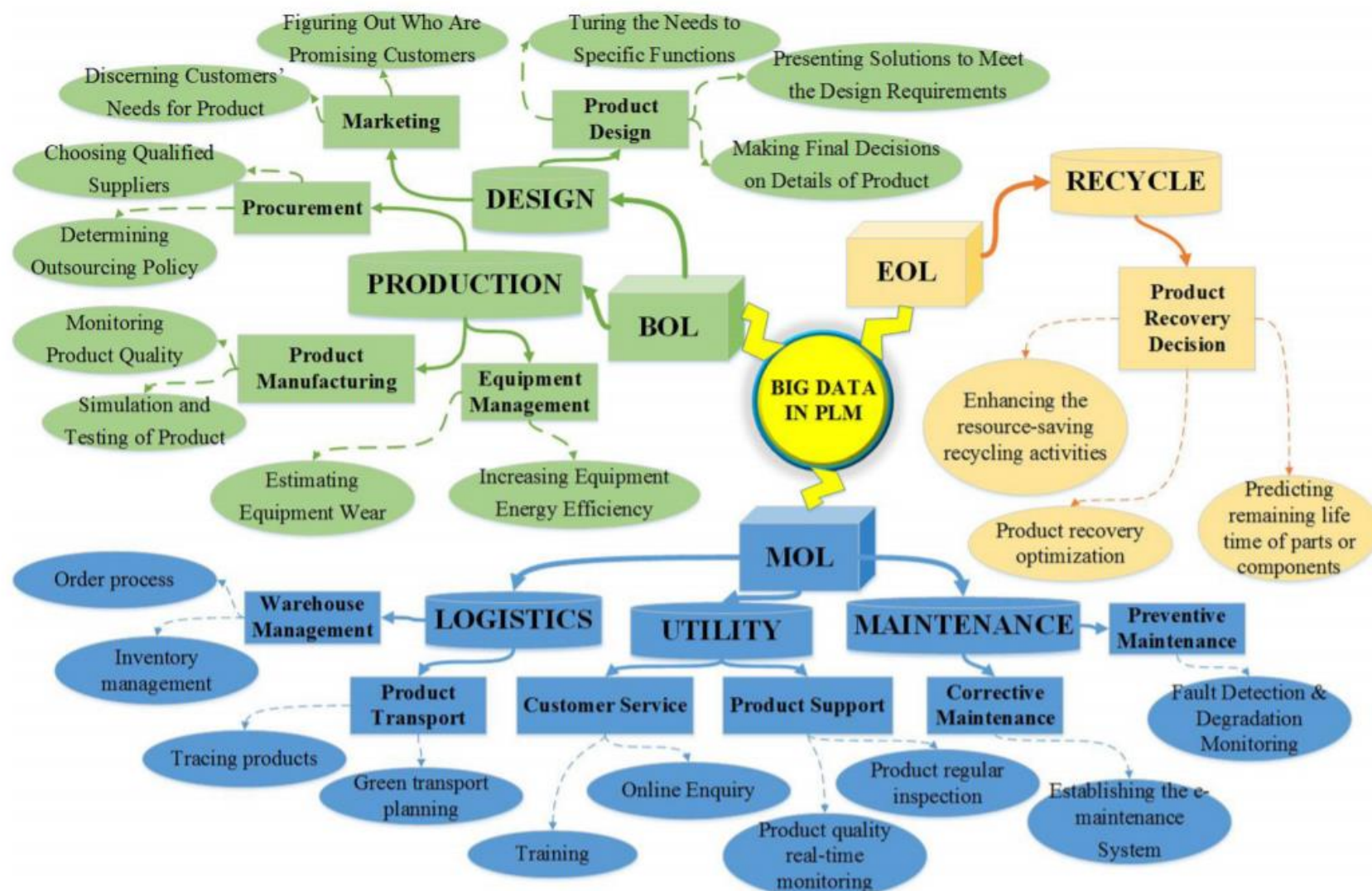


Figura 03: Framework com possíveis áreas de aplicação de *big data* no ciclo de vida de um produto.

Fonte: Li et al. (2015).

Em Erevelles et al. (2016), análise de dados de consumidores está no epicentro da revolução de *big data*, de modo que as informações em grande quantidade, variedade e velocidade são emitidas individualmente pelo comportamento e pelas vontades de cada consumidor. Para que as companhias possam aproveitar melhor esses dados em suas tomadas de decisão, o artigo propõe que os recursos da empresa criem valor das informações de *big data* nas esferas de precificação e inovação, e design de produtos. Seguindo nessa última esfera, Li et al. (2015) argumenta que análises de *big data* podem ser utilizadas em cada uma das fases do ciclo de vida de um produto. Para isso, os autores propõem um *framework* que sugere possíveis áreas com aplicação de *big data* no começo da vida (*BOL*), no meio da vida (*MOL*), e no fim da vida (*EOL*) de um produto. Na figura 03, pode-se observar a estrutura proposta pelos autores. Nela, há aplicações específicas em diversas dimensões, como o próprio design de produtos, marketing, gerenciamento de centros de distribuição, atendimento ao consumidor, e manutenção corretiva e preventiva.

Para Gunasekaran et al. (2017), a associação de *big data* com análise preditiva pode ter grandes implicações da cadeia de suprimentos e na performance organizacional de uma empresa. Para o artigo, a assimilação dessas informações na tomada de decisões de uma cadeia de suprimentos segue um processo de três passos: aceitação, rotinização e assimilação. Wang et al. (2016) também mostra que *big data* pode ser utilizada em diversas etapas da cadeia de suprimentos de uma empresa, como: planejamento de demanda, compras, produção, estoque, logística, análise estatística, simulação e otimização.

Para Murdoch e Detsky (2013), a aplicação de *big data* a indústria de *health care* é inevitável. Segundo eles, *big data* pode expandir a capacidade de gerar conhecimento no setor, e reduzir o custo de responder diversas perguntas clínicas. Além disso, pode ajudar na disseminação de informações, na tradução de iniciativas médicas pontuais em práticas clínicas, e pode ajudar a tornar os pacientes membros mais ativos do processo decisório em sua saúde. Já para Khoury e Ioannidis (2014), o bem-estar humano pode se beneficiar de dados em grande escala caso o ruído desses dados seja diminuído, pois em casos de informações pouco confiáveis, pode-se chegar a conclusões inválidas e a correlações falsas.

Zheng et al. (2013) propõe um método para inferir em tempo real informações a qualidade do ar através de uma série de diferentes informações da cidade de Pequim, na China, como meteorologia, tráfego de carros, mobilidade humana, estrutura de ruas e estradas, e pontos de interesse. Com o uso dessas informações, foi possível estimar a qualidade do ar com um nível muito maior de granularidade e precisão do que o permitido pela combinação entre os métodos tradicionais e os sensores instalados na cidade na época do estudo.

Em Zheng et al. (2016), a aplicação de *big data* ao transporte social traz oportunidades sem precedentes para a resolução de problemas para os quais abordagens tradicionais não são suficientes e para construir a nova geração de transportes inteligentes. Em Lv et al. (2015), pode-se observar outra aplicação de *big data* relacionada a transportes. Os autores propõem um método de prever o fluxo de tráfego por meio de *deep-learning*.

Também há estudos com aplicações de *big data* à mobilidade urbana e humana. Esses estudos utilizam dados de localização das pessoas em tempo real para capturar seus padrões de deslocamentos em dado centro urbano. Em Son et al. (2010), mostra-se que o grau de previsibilidade do comportamento dos deslocamentos humanos chega a 93% e que este índice é bastante independente das distâncias percorridas pelas pessoas, apesar da grande diferença dos padrões em si. Kang et al. (2011) também traz uma nova perspectiva às análises, propondo que os padrões dos deslocamentos humanos dentro de um ambiente são afetados por condições morfológicas das cidades, em especial a compactidade e o tamanho do ambiente em que estão inseridos. Com isso, conclui-se, por exemplo, que pessoas em cidades maiores ou menos densas tendem a ter a necessidade de se deslocar por distâncias maiores em seus dias a dias.

Destaca-se nessa frente de trabalho duas fontes principais de dados: (1) dados de telefonia celular, nos quais as trajetórias das pessoas são acompanhadas durante todo o seu dia; e (2) dados de deslocamentos de meios de transporte público ou privado individual, que acompanham os deslocamentos “porta a porta” de seus passageiros, podendo capturar as origens e os destinos destes movimentos.

González et al. (2008) argumenta que o conhecimento sobre as trajetórias das pessoas é pouco definido, apesar de ser importante em diversas esferas. O estudo, através do monitoramento da posição de cerca de 100.00 pessoas através de aparelhos celulares durante 6 meses, chegou à conclusão de que, na prática, os movimentos tendem a seguir padrões simples e reproduzíveis, em contraste com modelos utilizados, como modelos de *random walks*. Outros estudos foram feitos a partir de dados de telefones celulares, como em Calabrese et al. (2012), que busca dar um primeiro passo na criação de uma metodologia específica para aplicar esse tipo de dado em pesquisas de transporte, e novamente por González et al. (2017) que estuda o caso específico de Cingapura.

Em Peng et al. (2012), 1,58 milhões de dados de viagens de táxi em Shanghai, na China, são usados para definir os padrões de deslocamentos humanos na cidade. A partir das informações descobre-se que os deslocamentos dos usuários durante dias de semana são feitos com três objetivos principais: deslocamentos entre casa e trabalho, deslocamentos entre diferentes locais de trabalho, e outros tipos de movimentação, como deslocamentos para

atividades de lazer. Já em Liu et al. (2015), dados dessa mesma cidade são utilizados para definir a estrutura do local, incluindo a identificação de bairros e as regiões mais acessadas. Chu et al. (2014) faz procedimento semelhante para Shenzhen, também na China, coletando dados de trajetória de 21.360 táxis diferentes na cidade. O estudo transforma essas coordenadas nos correspondentes nomes de ruas para entender padrões e tendências de deslocamento na cidade.

2.2. Transporte individual de passageiros em grandes centros urbanos

Além de servir para definir a morfologia das cidades e os padrões de deslocamento dos usuários, os dados de transporte individual também têm outras aplicações mais específicas, como otimizar a utilização dos veículos, propor sistemas de visualização e realizar comparações entre os concorrentes públicos e privados.

Em Lee (2008), usa-se informações de um sistema telemático de táxis na ilha de Jeju-Do, na Coreia do Sul, que retorna a localização dos mesmos, junto a informação se o veículo possui passageiros ou não. Os dados são utilizados para mapear os tempos de espera dos táxis e os locais onde há a maior incidência de *pick-up* de passageiros.

Primeiramente, o autor verifica por inspeção visual quais são os locais com a maior incidência de busca de passageiros, cria clusters de localização e seleciona 10 *hot spots* para aprofundar a análise, que incluem zonas de interesse como a prefeitura da cidade e o aeroporto. Depois disso, estima a distribuição dos tempos de espera para buscar os passageiros em cada um dos diferentes pontos ao longo de um dia.

Dessa forma, o modelo permite sugerir aos taxistas quais são os locais preferidos para que eles obtenham sucesso ao buscar novos passageiros na cidade. Com isso, não apenas permite-se um aumento da rentabilidade dos profissionais ao diminuir os tempos de vacância dos veículos, como também se melhora o nível de serviço para os passageiros, que precisam esperar menos tempo para serem atendidos.

Em Cramer e Krueger (2016), os autores argumentam que, apesar da regulamentação dos táxis poder aumentar a segurança os consumidores e trazer outros benefícios, o limite de licenças imposto pelo governo também pode reduzir a eficiência da economia, dado que limita o número de carros com esse propósito em circulação e regulamenta o preço praticado pela categoria. Para comparar a eficiência dos táxis contra os serviços prestados por novas empresas de mobilidade, como Lyft e Uber, é feita uma análise sobre a capacidade utilizada em cada caso.

Os autores definem capacidade utilizada como o valor percentual de tempo ou de distância percorrida por um motorista em que há um passageiro no carro para pagar pela corrida realizada.

A análise foi feita comparando o táxi regular contra o Uber X, a categoria mais barata de transporte individual oferecida pela empresa em grandes centros urbanos dos Estados Unidos, como Boston, Los Angeles, Nova Iorque, São Francisco e Seattle. As informações de táxi foram coletadas de estudos ou portais de transparências específicos para cada cidade e as da Uber foram informadas pelo departamento de pesquisa da empresa. Concluiu-se através dessa análise que o Uber X tem uma utilização maior que o táxi nessas cidades, tanto em termos de horas quando de distância percorrida. Há quatro fatores na Uber que provavelmente colaboram para essa diferença. Eles têm: (1) alocação mais eficiente de passageiros com motoristas; (2) maior escala possibilitando alocações mais rápidas; (3) legislações ineficientes em táxis; e finalmente, (4) maior equilíbrio entre a oferta e a demanda ao longo de um dia, devido ao modelo de trabalho mais flexível e a tarifa dinâmica, que incentivam os motoristas a trabalhar em horários com maior procura pelo serviço.

Em Nie (2017), estudou-se o impacto que os novos entrantes tiveram nas corridas de táxi em Shenzhen, na China. Foram usados dados de mais de 2700 táxis, correspondentes a cerca de 18% do total da cidade, durante um período de 3 anos. Nessa base de dados, é possível identificar a localização dos táxis em diversos pontos georreferenciados emitidos cronologicamente (cerca de 90 segundos de intervalo entre os registros), que informam qual a corrida, qual o ponto de partida, o destino, e se o táxi estava vazio ou não.

Com essas informações, o autor buscou analisar como diversos aspectos da operação foram afetados, tanto em termos temporais quanto espaciais, com o objetivo de encontrar soluções de onde, como, e quando, os taxistas poderiam competir de maneira mais efetiva. Estes aspectos analisados foram: (1) número de viagens médio por hora; (2) distância e tempo médios viajados com e sem ocupação de passageiros; (3) produtividade, medida pela velocidade média e pela capacidade utilizada; e, por fim, (4) distribuição espacial das corridas relativas à densidade populacional e horários de pico.

No estudo, descobre-se que houve um choque bastante forte nos táxis devido ao aumento de competição, com uma redução grande na quantidade de corridas feitas na cidade (cerca de 25%). Além disso, é averiguado que esse impacto foi bastante abrupto, ocorrendo em menos de um período de 12 meses, e que já está acontecendo uma estabilização dos serviços de táxi. Ainda mais, conclui-se que os táxis competem melhor em zonas com maior densidade populacional e em horários de pico, que tornam o *e-hailing*, como é chamado o processo de

pedir um carro pelo aplicativo, menos atrativo, e que táxis que também utilizam esse processo tendem a ter um aumento da sua utilização. Com isso, o autor argumenta que os dois serviços precisam coexistir. Assim como os táxis não conseguem tomar de volta sua força de mercado, a possibilidade de pedir um táxi na rua, que não pode ser replicada pelos concorrentes, ainda é bastante demandada e torna o serviço único.

Em McKenzie e Baéz (2016), utiliza-se dados de corridas de táxi e Uber na cidade de Nova Iorque para detectar “eventos” que fogem do padrão na rotina da cidade e busca diferenças entre os eventos percebidos nas bases de cada um dos dois tipos de transporte. Os dados utilizados para os táxis foram obtidos em informações divulgadas pela *TLC*.

Primeiramente, encontrou-se os eventos em cada uma das bases. Para isso, calculou-se a média de corridas realizadas e o desvio padrão para cada divisão de área da cidade informada no censo de Nova Iorque, hora a hora. Considerou-se um evento detectado toda vez que o número de *pick-ups* em dada localização superou em três vezes o desvio padrão calculado.

Nessa análise, foram encontrados eventos em ambas as plataformas. Entretanto, ao buscar os eventos de Uber correspondentes na base de táxis, descobriu-se que a grande maioria dos eventos era exclusiva de cada plataforma. De um total de 485 eventos detectados para os Ubers e 2671 para os táxis, apenas 17 ocorreram nas duas plataformas, mesmo adotando uma margem de 2 horas de diferença para buscar a correspondência.

Além disso, por meio de uma verificação manual dos eventos, o estudo segmentou os públicos de cada tipo de evento em diferentes categorias. Os participantes dos eventos exclusivos de Uber têm uma tendência a serem mais jovens e mais conectados com tecnologia, enquanto os de táxi parecem ter características de pessoas mais velhas. Ocorrências detectadas em ambas as bases foram grandes espetáculos, que possuem público numeroso e com variação ampla de idade.

Em Ferreira et al. (2013), dados da *TLC* das corridas de *Yellow Taxi* em Nova Iorque são utilizados para criar um modelo no qual os usuários podem buscar *queries* com os dados pretendidos, permitindo-nos estudar visualmente análises de origem-destino com os parâmetros selecionados. Outra proposta de procedimento com a mesma finalidade é realizada por Doraiswamy et al. (2018), utilizando-se o software *Urbane* para criar uma visualização interativa com os desejos dos usuários.

Ainda no artigo de Ferreira et al. (2013), os autores consideram os dados de táxi como um sensor da vida da cidade. Segundo eles, cerca de 13.000 táxis fazem em média 500.000 corridas por dia para por volta de 1 milhão de passageiros.

Segundo eles, lidar com a base de dados de táxis de Nova Iorque é bastante desafiador devido ao grande volume de dados. Muitas vezes, a solução empregada pelos cientistas é utilizar apenas uma fração dos dados ao invés de capturar tudo que os registros podem trazer. Ferramentas amplamente utilizadas, como R, MatLab, Stata, ArcGIS e Excel não são capazes de lidar com todos os números. Ainda mais, esse grande volume transforma o processo de análise em algo tedioso e que consome muito tempo.

Para ajudar a suprir essa necessidade de análise, criou-se um sistema chamado TaxiVis que permite ao operador do sistema fazer *queries* nas quais o usuário escolhe as faixas de tempo que serão analisadas e seleciona visualmente as regiões geográficas para as quais quer acessar as informações, de modo que o próprio usuário diretamente manipula os dados.

A arquitetura do sistema obedece aos parâmetros representados na figura 04.

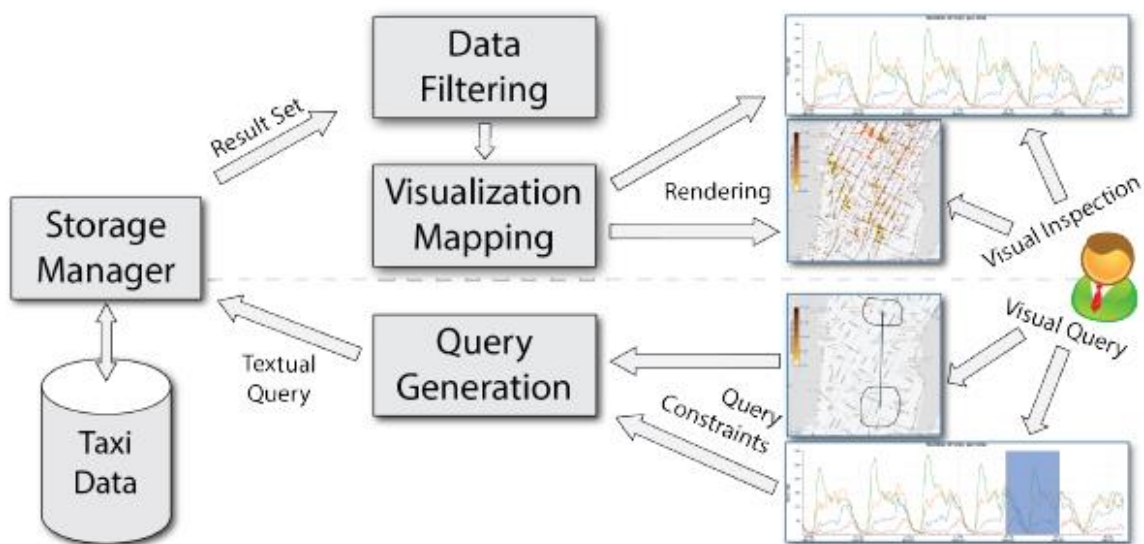


Figura 04: Parâmetros da arquitetura do sistema TaxiVis.

Fonte: Ferreira et al. (2013).

Para o usuário, o sistema aparece como pode ser observado na figura 05. Na região “A” demarcada em vermelho, o usuário escolhe a data de início e término da visualização dos dados e o intervalo de tempo em que será exposto o resultado. Na região “B”, pode-se selecionar por meio do mapa as regiões que deverão aparecer na análise. Na “C”, há uma barra de ferramentas na qual escolhe-se alguns dos parâmetros que serão utilizados, como o conjunto de dados a ser analisado (*pick-up*, *drop-off* ou ambos) e possibilidades de exportação das informações. Na parte “D”, plota-se os dados referentes aos parâmetros selecionados nos itens anteriores. Pode-se também escolher entre diferentes formas de visualização, como séries temporais e histogramas.

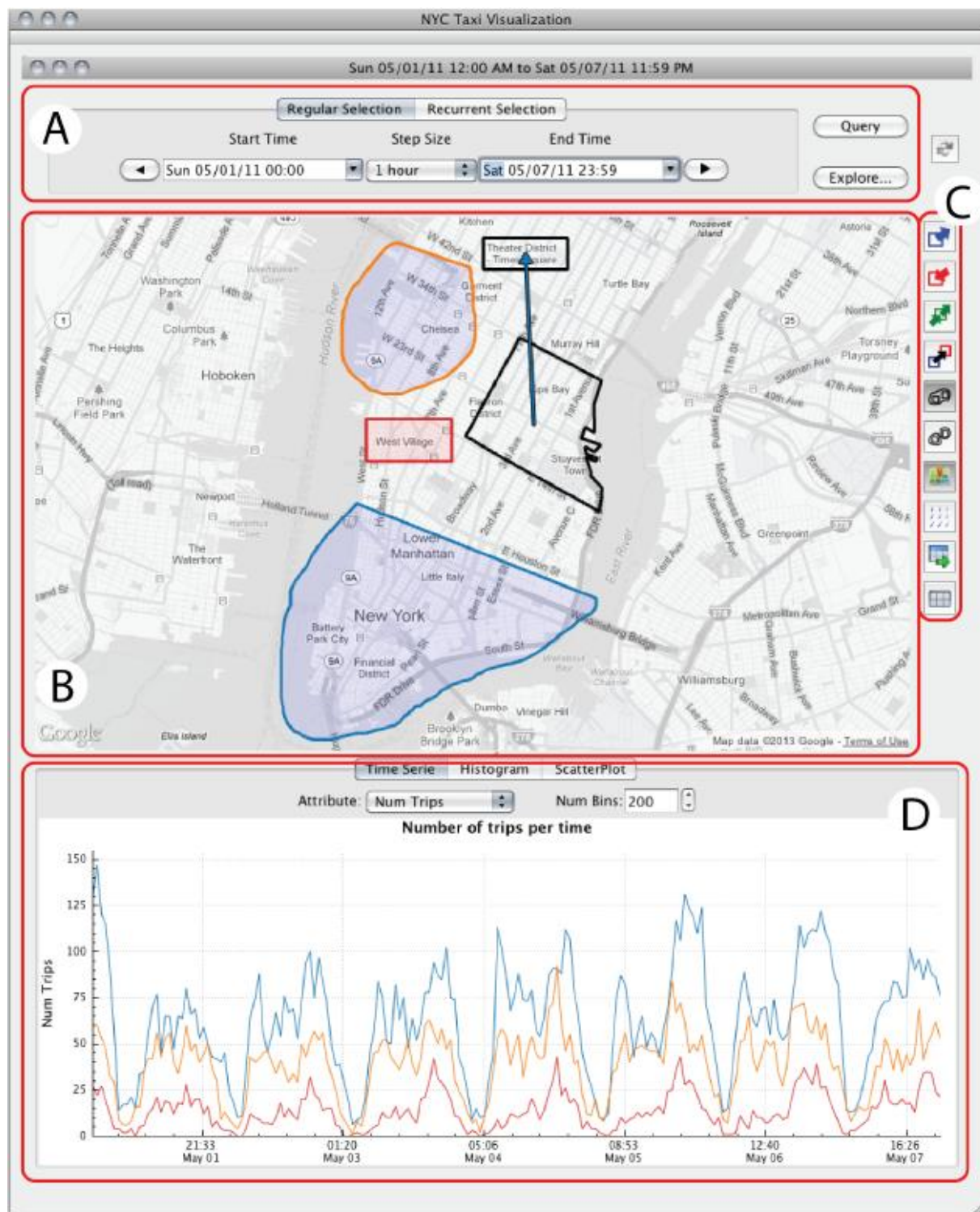


Figura 05: Visualização para o usuário da seleção de parâmetros analisados e resultados do TaxiVis.
Fonte: Ferreira et al. (2013).

Para mostrar como a ferramenta é utilizada e como ela pode gerar informações relevantes, o artigo traz algumas análises para exemplificar análises que podem ser realizadas pela ferramenta.

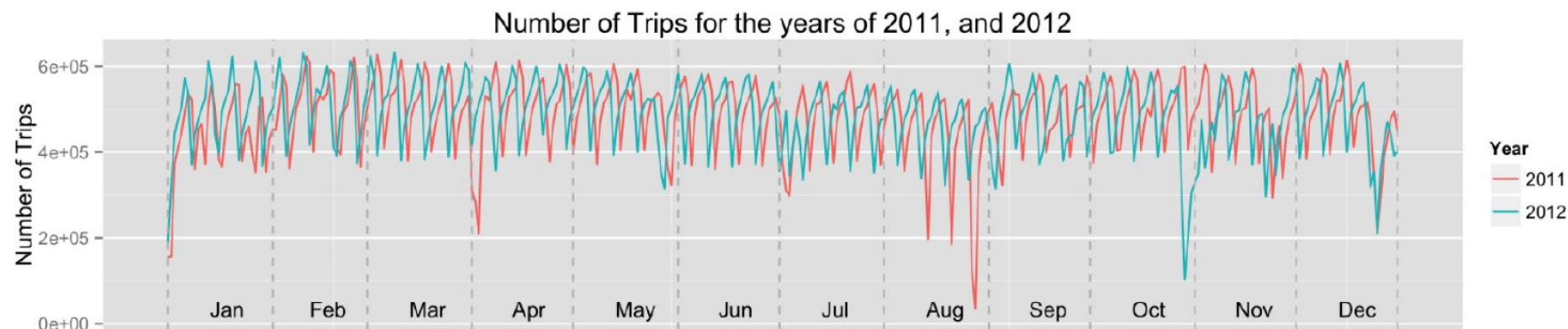


Figura 06: Número de corridas de *Yellow Taxi* realizadas na cidade em 2011 e 2012.
Fonte: Ferreira et al. (2013).



Figura 07: Número de *pick-ups* e *drop-offs* realizados ao longo da primeira semana de maio de 2011 para diferentes bairros do *borough* de Manhattan.
Fonte: Ferreira et al. (2013).

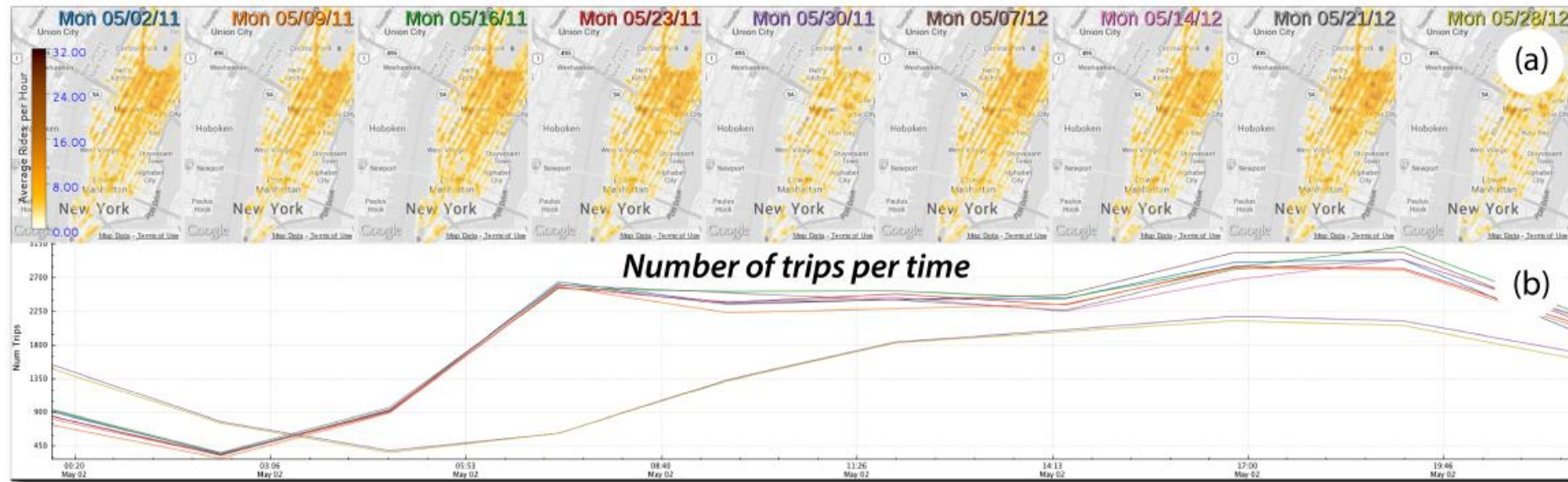


Figura 08: Distribuição espacial e por horário da quantidade de corridas realizadas em Manhattan em diferentes segundas feiras para os anos analisados.
Fonte: Ferreira et al. (2013).

Por meio da figura 06, com corte para todas as corridas de táxi realizadas na cidade diariamente entre 2011 e 2012, os autores mostram que o este número varia ano a ano, mas que existe uma forte sazonalidade entre os dias e as épocas do ano. Além disso, há datas festivas, como o Natal, o Ano Novo e o Dia de Ação de Graças em que há uma queda substancial no número de corridas que pode ser observada nos dois intervalos. Ademais, ressalta-se que eventos esporádicos podem ser capturados pela variação no número de corridas, como ocorreu com a chegada dos furacões Irene e Sandy na cidade de Nova Iorque, representadas nas quedas bruscas no número de viagens realizadas no final dos meses de agosto de 2011 e outubro de 2012, respectivamente.

Na figura 07, é realizada uma análise para comparar a quantidade de *pick-ups* e *drop-offs* de táxis em diferentes regiões de Manhattan na primeira semana de maio de 2011. Pode-se observar que existe uma repetição entre os padrões nos mesmos horários para diferentes dias, especialmente durante a semana. Além do mais, os padrões entre *pick-ups* e *drop-offs* são diferentes e variam bastante entre os diferentes bairros analisados.

Já na figura 08, mostra-se em “(a)” um mapa de calor com a atividade das corridas de táxi de Táxi em Manhattan para diferentes segundas feiras ao longo de 2011 e 2012 e em “(b)” um gráfico com a quantidade de corridas nesses respectivos dias ao longo da duração de um dia. Pode-se perceber que há dois casos em que o número de corridas é visivelmente menor, que são 30/05/2011 e 28/05/2012.

3. METODOLOGIA

3.1. Metodologia de pesquisa

3.1.1. Apresentação da metodologia escolhida

Segundo Bonomo (2007), uma pesquisa deve obedecer a uma série de critérios, que por sua vez seguem uma classificação baseada nos objetivos gerais da mesma. Os principais critérios são a (1) a finalidade, (2) os objetivos, (3) os procedimentos, (4) a natureza e (5) o local de realização.

O presente trabalho segue a metodologia de pesquisa apresentada na figura 09.

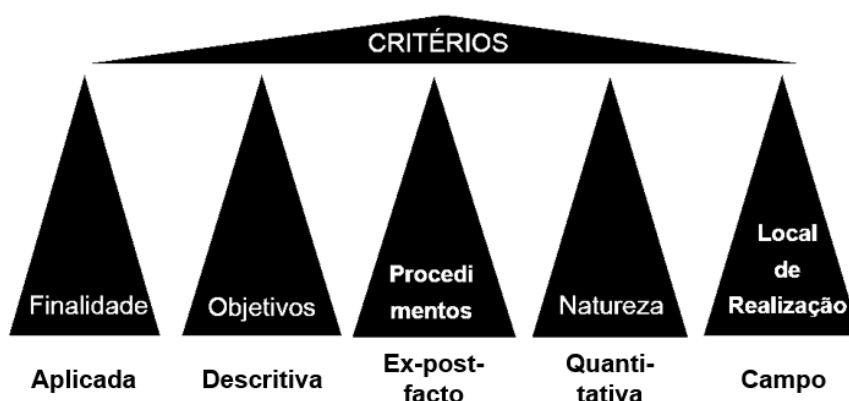


Figura 09: Critérios de pesquisa do presente trabalho.
Fonte: elaborado pelo autor com base em Bonomo (2007).

Segundo este autor, a finalidade de uma pesquisa pode ser básica ou aplicada. A pesquisa básica é aquela que representa o desenvolvimento de teorias e gera conhecimentos para serem utilizados pelas ciências aplicadas. As ciências aplicadas são as que utilizam as informações geradas pelas ciências básicas para gerar produtos, processos ou conhecimento com finalidade imediata de acordo com às necessidades humanas. Este estudo segue a finalidade aplicada pois busca entender na prática como foram as variações entre os padrões de uso dos táxis na cidade de Nova Iorque e pode ser usado por diversos segmentos, desde órgãos reguladores que buscam entender os efeitos em mobilidade urbana da entrada de serviços de corrida compartilhados como a Uber em diferentes cidades até grandes companhias de táxi para buscar desenvolver melhor suas reações competitivas.

O mesmo autor apresenta ainda o objetivo de uma pesquisa como exploratório, descritivo ou explicativo. A pesquisa exploratória é tida como uma primeira aproximação para um tema, de modo a ganhar maior familiaridade com o mesmo. A descritiva é aquela que levanta características conhecidas do fato estudado e é feita por meio de levantamentos ou observações sistemáticas do mesmo. A explicativa visa explicar ou criar teorias a respeito de

um fenômeno e a aprofundar o conhecimento sobre a realidade. O presente estudo é classificado como descritivo pois descreve como eram os padrões de deslocamento dos táxis no mercado de Nova Iorque antes e depois da entrada da Uber, por meio de visualizações de dados já existentes que representam esse fenômeno, sem se preocupar em explicar como essas mudanças ocorreram.

Ainda segundo Bonomo (2007), há diversos procedimentos diferentes de coleta de dados que podem ser utilizados em um projeto de pesquisa, como a pesquisa experimental, que realiza experimentos científicos; o *ex-post-facto*, no qual o investigador não tem controle direto sobre as variáveis e inferências são feitas por observação direta; o levantamento, em que há a interrogação direta do público alvo do estudo; o estudo de caso, que busca aprofundar em um caso específico; a pesquisa bibliográfica, realizada a partir de trabalhos anteriores publicados; e a pesquisa documental, na qual é dada uma ênfase em informações que ainda não foram publicadas no meio acadêmico. O procedimento adotado para este estudo foi o *ex-post-facto* pois os dados coletados se referem a corridas de táxi que já ocorreram e o autor não possui controle direto nem sobre as variáveis e nem as informações colhidas.

A natureza da pesquisa pode ser qualitativa ou quantitativa. Segundo Miguel et al. (2012), a abordagem qualitativa possui interesse na perspectiva do fenômeno estudado, enquanto a quantitativa tem esforços empenhados na mensuração das variáveis da pesquisa. Dessa maneira, o estudo presente segue a abordagem quantitativa por meio de análise de dados.

A fonte de dados pode ser de campo, laboratório ou pesquisa bibliográfica. A de campo estuda as informações onde elas ocorrem, por meio de coleta de dados e observações dos fenômenos estudados *in natura*. O laboratório consiste na realização de experimentos que permitem interferir artificialmente na produção dos fatos ou artificializar o ambiente para que seja produzido de maneira adequada. A pesquisa bibliográfica consiste em desvendar, organizar e analisar os principais fatos já estudados sobre o referente tema. Este estudo segue a abordagem de campo pois as informações utilizadas são reportadas e recolhidas dos próprios táxis de Nova Iorque, de modo que são extraídas diretamente da realidade do objeto de estudo, sem a interferência dos pesquisadores.

3.2. Metodologia do projeto

3.2.1. Fonte de dados

O estudo foi desenvolvido com base nos dados de corridas de táxi e *FHV* disponibilizados pelo *TLC*. O presente estudo é focado nos *Yellow Taxis*, que serão chamados de táxi para efeitos de simplificação. Esses táxis comuns podem realizar corridas em todas as

regiões da cidade, e os *FHVs* são veículos de livre circulação que realizam transporte de passageiros. A Uber é uma empresa que atua no mercado de *FHVs* e suas corridas podem ser identificadas na série de dados fornecida pela *TLC*.

Segundo a definição do próprio órgão, a *TLC* é uma agência governamental criada em 1971 que é a responsável pelo licenciamento e a regulação dos táxis e *FHVs* em Nova Iorque (*TLC*, 2019). Os dados divulgados pela mesma correspondem a todas as corridas finalizadas por táxis e *FHVs* licenciados pela instituição. Para fins de coleta, cada táxi é equipado com medidores eletrônicos que reportam as informações automaticamente para a agência e as informações referentes às corridas de *FHV* são enviadas diretamente pelas empresas responsáveis pelo transporte.

Os dados referentes a cada mês foram copiados individualmente e depois tratados formando grandes bases de dados unificadas. As bases de dados em formato *.csv*, entre 2009 e 2018 para os *Yellow Taxis*, totalizam cerca de 225 *gigabytes* de dados; e entre 2015 e 2018 para os *FHVs*, cerca de 32 *gigabytes*.

Além disso, também foi utilizada a base com os *shapefiles* das zonas de táxi da cidade fornecidas pela mesma agência, no formato *.shp*. Esse arquivo possui a delimitação geográfica de cada zona, a descrição do local, a sua identificação e *borough* ao qual ela pertence.

3.2.1.1. Descrição das variáveis de estudo

As variáveis do estudo foram obtidas por meio dos dados fornecidos pela *TLC*. Os dados de táxis e *FHVs* são diferentes entre si e variam ano a ano. Nas tabelas 1 e 2, há um quadro-resumo para cada caso, com as principais informações divulgadas por período. No caso dos táxis, além das informações representadas, há outras que não compõe a base utilizada neste estudo, como meios de pagamentos, taxas, indicação se o aparelho eletrônico que envia as informações estava conectado à internet no momento da finalização da corrida, entre outras. Cabe ressaltar que nem todas as análises preliminares aparecem neste documento, de modo que algumas das variáveis presentes na base utilizada não aparecem nos resultados.

Tabela 01: Quadro-resumo com as informações divulgadas ano a ano para táxis que foram analisadas no estudo.

Ano	Data e horário de <i>pick-up</i>	Data e horário de <i>drop-off</i>	Número de passageiros	Distância percorrida	Valor da corrida	Latitude e longitude de <i>pick-up</i>	Latitude e longitude de <i>drop-off</i>	Zona de táxi de <i>pick-up</i>	Zona de táxi de <i>drop-off</i>
2009	✓	✓	✓	✓	✓	✓	✓		
2010	✓	✓	✓	✓	✓	✓	✓		
2011	✓	✓	✓	✓	✓	✓	✓		
2012	✓	✓	✓	✓	✓	✓	✓		
2013	✓	✓	✓	✓	✓	✓	✓		
2014	✓	✓	✓	✓	✓	✓	✓		
2015	✓	✓	✓	✓	✓	✓	✓		
2016	✓	✓	✓	✓	✓	✓	✓		
2017	✓	✓	✓	✓	✓			✓	✓
2018	✓	✓	✓	✓	✓			✓	✓

Fonte: elaborado pelo autor com base nas informações divulgadas pela *TLC*.

Tabela 02: Quadro-resumo com as informações divulgadas ano a ano para *FHVs* que foram analisadas no estudo.

Ano	Origem dos dados	Data, horário e zona de táxi de <i>pick-up</i>	Data, horário e zona de táxi de <i>drop-off</i>	Informação de corrida compartilhada
2015	✓	✓		
2016	✓	✓		
2017 (jan-mai)	✓	✓		
2017 (jun)	✓	✓	✓	
2017 (jul-dez)	✓	✓	✓	✓
2018	✓	✓	✓	✓

Fonte: elaborado pelo autor com base nas informações divulgadas pela *TLC*.

Além das variáveis fornecidas, algumas outras variáveis foram construídas para realizar análises adicionais. Elas são: (1) dia da semana (obtido por meio das datas das corridas, feito sempre com base no *timestamp* do *pick-up*), (2) zonas de táxi de *pick-up* e *drop-off* para corridas de táxi em 2009 e 2013 (a partir das latitudes e longitudes cruzadas com o arquivo *.shp* com os *shapefiles* das zonas definidas pela *TLC*, para tornar os dados mais recentes comparáveis com os antigos) e (3) empresa responsável pela corrida de *FHV* (para poder isolar os dados das corridas da Uber, obtido por meio da coluna de origem dos dados).

3.2.1.2. Restrições dos dados

As bases de dados utilizadas são apenas as disponíveis no portal da *TLC* na internet e os dados disponíveis variam ano a ano. Os arquivos de táxi estão disponíveis entre 2009 e 2018 e os de *FHVs* entre 2015 e 2018. Entretanto, em alguns casos os dados copiados diretamente do site estavam corrompidos para os táxis, mais especificamente para os meses de fevereiro e março de 2010, março de 2015 e agosto a dezembro de 2016. Nesses casos, as figuras para os anos completos foram estimadas por meio de interpolação simples.

Além disso, conforme documentado na seção anterior (3.2.1.1.), as informações divulgadas pela *TLC* a respeito de táxis e *FHVs* possuem níveis de detalhamento diferentes, que podem também apresentar variações entre os meses de estudo.

No caso dos táxis, o caso mais relevante foi o das informações de latitude e longitude de *pick-up* e *drop-off* que foram substituídas a partir de 2017 pela identificação da região de início e término do transporte segundo as zonas de táxi de Nova Iorque definidas pela *TLC*.

Para os *FHVs*, há um elevado nível de restrição de informação, incluindo a distância percorrida e dados referentes ao horário e localização de *drop-off* dos passageiros, que só foram disponibilizados a partir de junho de 2017.

Cabe também ressaltar que, como foi encontrado tipicamente em outros problemas envolvendo *big data* (Liu et al., 2012), há muitas vezes linhas de informação nas bases que são pouco condizentes, como corridas de táxi de 99999999 quilômetros ou latitudes e longitudes representadas fora da cidade de Nova Iorque. Na maioria desses casos, a quantidade de *outliers* não foi muito expressiva (menos de 1% da base), portanto optou-se por não realizar nenhum tratamento e preservar os dados com o mínimo possível de tratamento, de modo a facilitar a replicação das análises deste estudo no futuro. O caso mais expressivo foi o de corridas cuja data de *pick-up* não correspondia com o ano para o qual a base foi extraída. Nestes casos, os registros correspondentes foram removidos (somente para táxis).

3.2.2. Ferramentas de análise utilizadas

Para analisar os dados, foi necessário valer-se de diferentes ferramentas que dão suporte ao grande volume de informações manipulado. As principais ferramentas e seus usos foram o Alteryx Designer 11.0 (Alteryx, 2018), utilizada para preparar os dados, realizar análises avançadas e sumarizar os grandes volumes de dados em bases mais enxutas; o Tableau Desktop 2019.2 (Tableau Software, 2019), empregado para visualizações mais avançadas das informações; e o Microsoft Excel 2016 (Microsoft, 2016a), para obter informações simples e consolidadas a partir das bases produzidas no Alteryx. Os gráficos foram feitos no software Mekko Graphics (Mekko Graphics, 2018), embutido no Microsoft PowerPoint (Microsoft, 2016b).

O Alteryx e o Tableau, por serem ferramentas novas e de grande importância para o trabalho, tiveram seu uso descrito em maior detalhe nas seções abaixo.

3.2.2.1. Alteryx Designer

O Alteryx é um *software* utilizado para a preparação e análise avançada de dados. Dentro da plataforma, é possível juntar diversas informações de fontes diferentes, preparar e tratar os dados que serão utilizados, e realizar análises avançadas de diversas categorias, como análises descritivas e preditivas, análises de segmentação, análises demográficas e de comportamento, e análises espaciais.

Com uma abordagem de “*drag-and-drop*”, o software permite que o usuário realize operações muitas vezes complexas em termos de códigos em linguagens de programação apenas com poucos cliques no computador. Além disso, criam-se processos que podem ser facilmente repetidos para diferentes fontes de dados, inclusive com volumes muito grandes de informação.

Na figura 10, observa-se a tela principal do Alteryx Designer, na qual o usuário cria um fluxo com os dados que executa as operações por ele desejadas. Nesse fluxo, o usuário coloca três fontes de dados diferentes dentro de “*Input Data*”, depois prepara os dados em “*Data Preparation*”. Em uma próxima etapa, junta as informações das diferentes fontes em “*Data Blending*” e realiza uma análise geográfica em “*Geographic Analysis*”, que é encaminhada para uma análise preditiva em “*Predictive Analytics*” e para mecanismos de exportação em “*Share Analyses / Data*”.

Os comandos são feitos selecionando os ícones que estão na parte de cima da tela (os ícones que podem ser visualizados são marrons, dentro da seção de análise preditiva) e os arrastando para a parte de baixo, na área branca na qual se encontram os ícones do exemplo. Depois, deve-se criar elos entre os diferentes ícones, que estabelecem a ordem na qual os processos devem ocorrer com os dados selecionados. Finalmente, basta o usuário apertar para rodar o fluxo construído, já com as funções de exportação definidas.

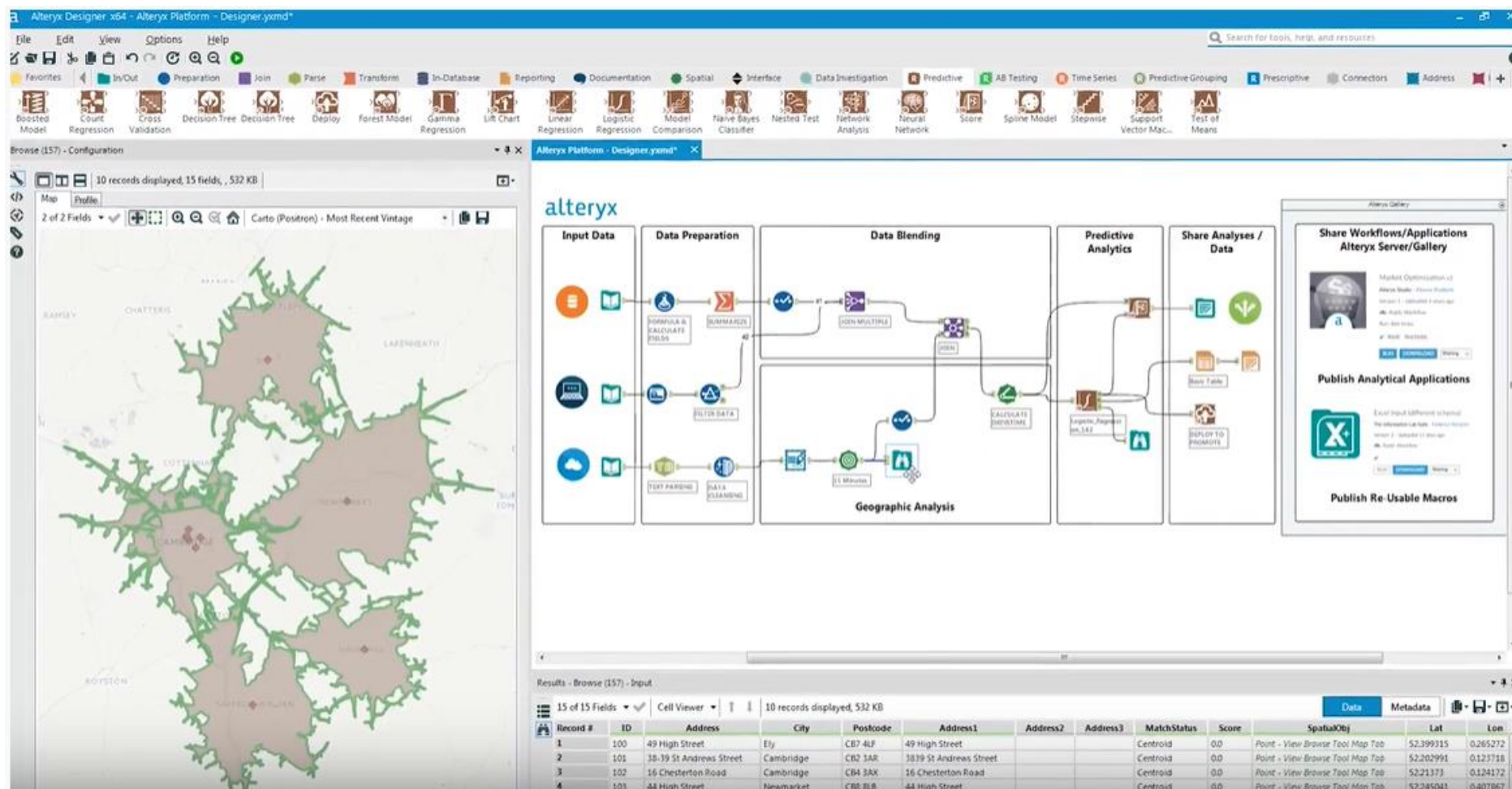


Figura 10: Exemplo de fluxo do Alteryx, mostrando o tratamento e a junção de três bases de dados.
 Fonte: tutorial de apresentação do Alteryx divulgado pela própria companhia.

3.2.2.2. Tableau Desktop

O Tableau é um *software* usado principalmente para a visualização de dados e para a construção de *dashboards* interativos.

Na figura 11, mostra-se uma base de dados já dentro do programa. Do lado esquerdo, aparecem todas as colunas presentes no arquivo, classificadas em dimensões ou medidas. Arrastam-se esses atributos para serem visualizados nas áreas desejadas pelo operador do *software*. Eles podem ser arrastados para a seção de linhas, colunas, páginas, filtros ou marcas. A seção de marcas é a mais completa, pois nela pode-se representar diferentes valores dentro do mesmo *set* de dados, sendo possivelmente representados como cor, tamanho, legenda, detalhe ou *tooltip*. No lado direito, abaixo de “*show me*”, aparecem as diferentes visualizações sugeridas para os dados, que são permitidas segundo as dimensões e medidas que já foram selecionadas pelo usuário.

Uma forma de visualização específica para a qual o Tableau é bastante forte é a visualização de mapas, que consegue ler e representar com facilidade os *shapefiles .shp* das zonas de táxi do *TLC*, que é bastante explorada neste trabalho, como no mapa com as zonas de táxi por *borough* visualizado na figura 02.

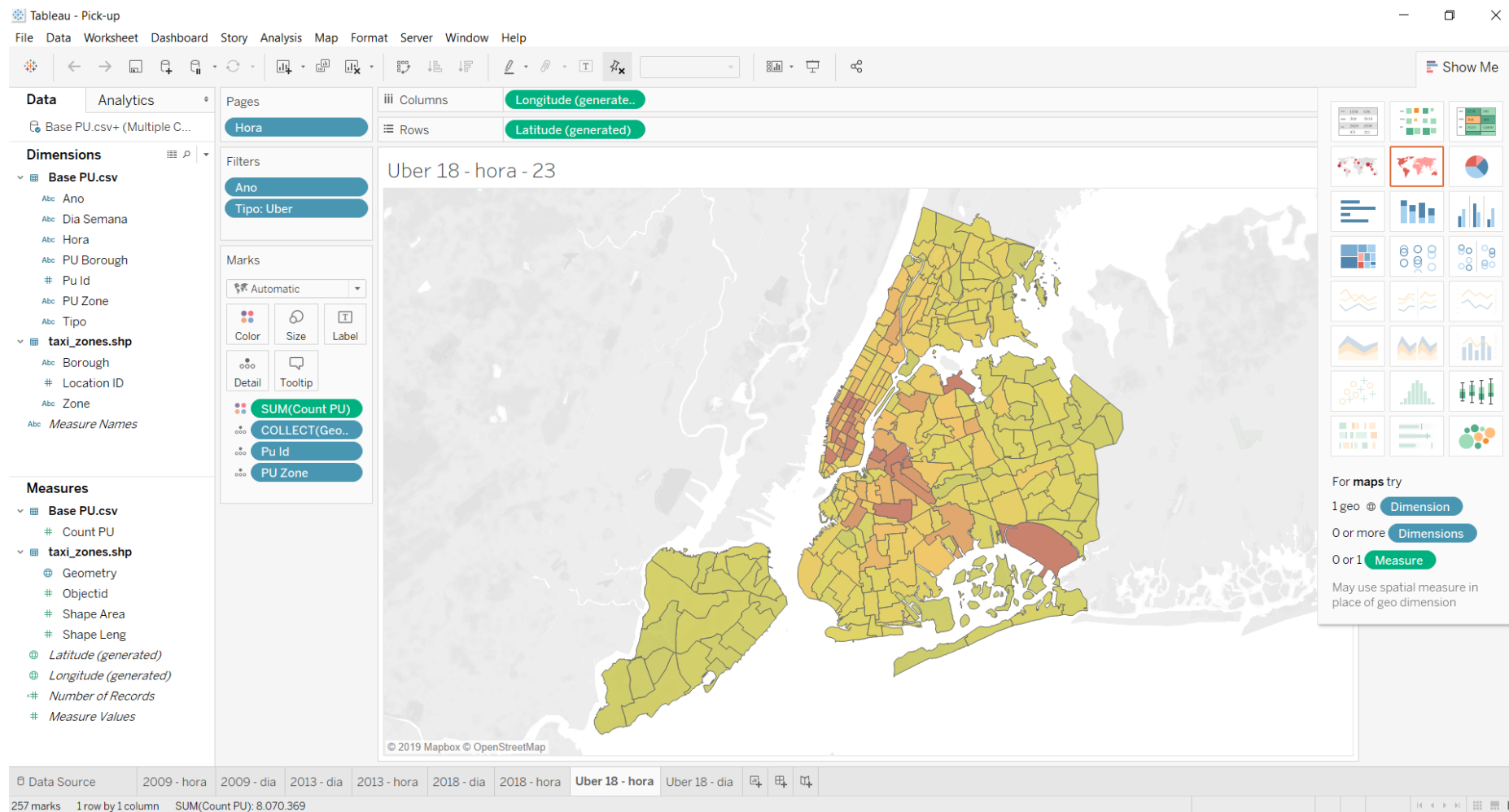


Figura 11: Tela do Tableau com visualização geográfica de dados da cidade de Nova Iorque.
Fonte: elaborado pelo autor.

4. PROJETO

4.1. Preparação das bases de dados

Antes das análises, foi feito um tratamento prévio dos dados para deixar as bases prontas para uso no Alteryx. Tanto para táxis quanto para *FHVs*, foram criadas bases adicionando os campos de (1) “Mês”, com o mês e o ano do *pick-up* da corrida; (2) “Dia_Semana”, com o dia da semana em que ocorreu o *pick-up* da corrida; e (3) “Tipo”, que marca se a corrida foi feita por um táxi ou pela operadora que executou a corrida em caso de *FHV*. Em uma quantidade significativa de registros, não há nenhuma operadora marcada para os *FHVs*. Ao entrar em contato com a *TLC*, foi dito que essas corridas realmente não foram feitas por nenhuma das operadoras que são identificadas (Uber, Lyft, Juno e Via). Dessa maneira, como o caso analisado foi somente o da Uber, esses registros foram descartados.

Para os dados de táxi, como os dados anteriores à 2016 tem as coordenadas de *pick-up* e *drop-off* registradas ao invés das zonas de táxi, selecionou-se dois anos para fazer um tratamento adicional, de modo a realizar um cruzamento das informações de latitude e longitude com as zonas espaciais fornecidas pela *TLC* e incluir uma coluna de localização equivalente com a zona de táxi nos dados. Os anos escolhidos foram 2009 e 2013, dois anos com dados sem falta de registros, com distância temporal entre si, e nos quais a entrada da Uber em Nova Iorque ainda não podia ser percebida com diferenças significativas em termos da movimentação dos táxis. Desta forma, é possível analisar se já haviam mudanças de movimentação antes da entrada da Uber na cidade, e depois comparar com os dados de 2018, nos quais já há uma grande diferença no número de corridas dado o novo entrante.

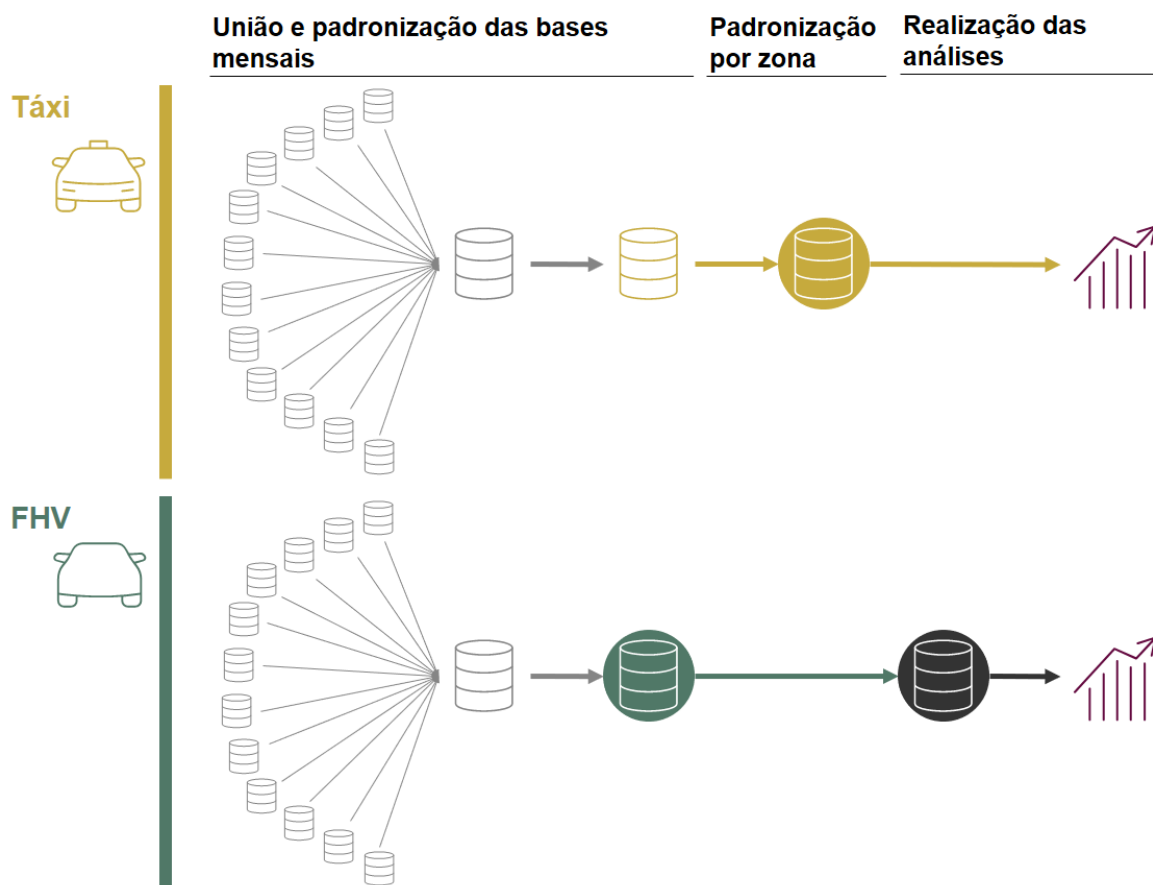


Figura 12: quadro ilustrativo com o processo de preparação dos dados.
Fonte: elaborado pelo autor.

Na figura 12, temos um resumo do processo de preparação dos dados. Primeiro são formadas bases anuais com a união dos 12 meses divulgados pela *TLC* para cada ano, tanto para táxi quanto para *FHV*. Depois, tem-se um passo no qual são adicionadas colunas contendo os campos adicionais descritos acima, padronizando os dados e formando uma base intermediária para os táxis e uma base final para os *FHVs*. Em seguida, a base de táxi passa por uma padronização por zonas de táxi (para 2009 e 2013), incluindo essa informação, que foi obtida pelo cruzamento das latitudes e longitudes das corridas com o *shapefile* da *TLC*. Nesse momento, as duas bases já estão prontas para serem analisadas, sendo que, antes cada uma das análises, é necessário filtrar a base de *FHV* para obter somente os dados da Uber.

Esse processo foi repetido para todos os anos com informações divulgadas pela agência estadunidense.

4.1.1. Dados de táxi

O tratamento realizado consistiu em unir as bases de dados mês a mês para cada um dos anos no Alteryx, criar as colunas necessárias (“Mês”, “Dia_Semana” e “Tipo”) e exportar o

resultado como uma base nativa do programa, que ocupa menos espaço e pode ser executada com maior velocidade.

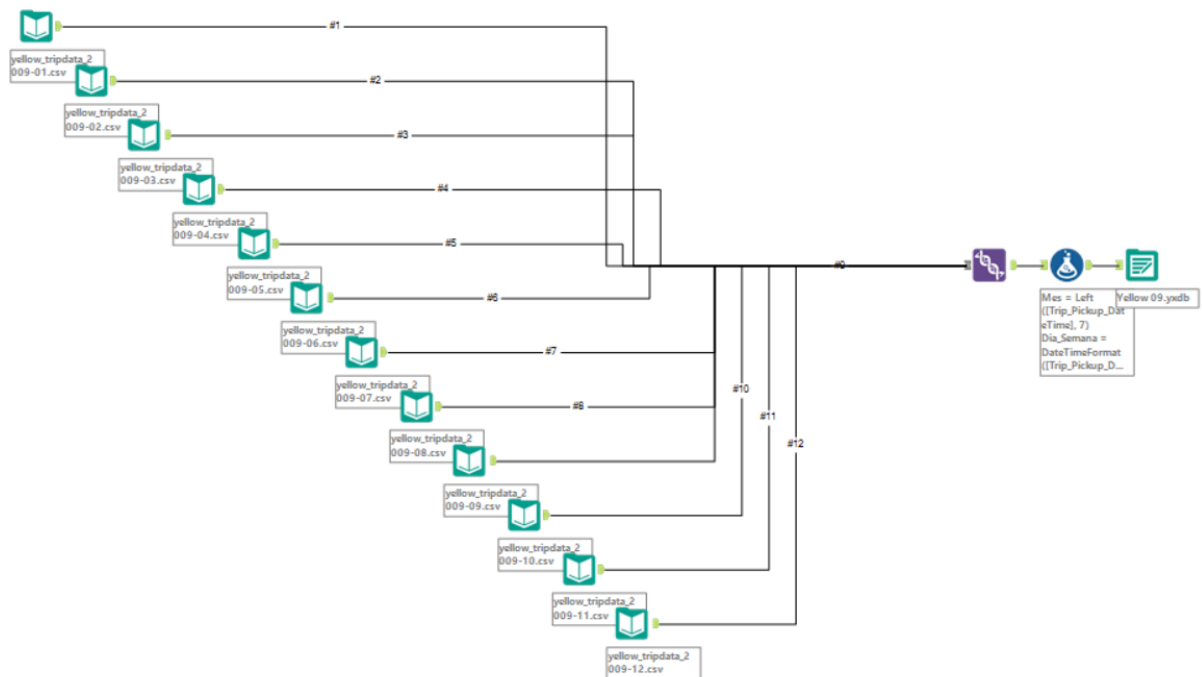


Figura 13: Fluxo do Alteryx que representa o tratamento inicial dos dados de táxi para o ano de 2009.
Fonte: elaborada pelo autor.

Na figura 13, tem-se do lado esquerdo cada uma das bases mensais de 2009 colocadas por meio de ícones verdes de “Input Data” no sistema. Depois, no lado direito, os dados são unidos com a ferramenta roxa de “Union”, formando uma base única com todas as informações. Em seguida, são criadas as colunas adicionais por um ícone azul de “Formula”, completando a base que será utilizada. Para isso, foi designada uma fórmula específica que transforma as informações que já estavam presentes nos dados em cada uma das colunas necessárias. Por fim, os dados são salvos com o comando verde de “Output Data”, já em um formato de arquivo nativo do Alteryx, aumentando a velocidade das análises que foram realizadas com estas informações.

Esse processo foi repetido para todos os anos entre 2009 e 2018, utilizando os meses com dados disponíveis, gerando bases anuais que serão utilizadas nas análises.

4.1.2. Dados de FHV

Para os dados de FHV, foi realizado um processo semelhante, de unir todas as bases mensais para cada ano e adicionar as mesmas colunas. Foi adicionado ao processo a ferramenta de roxa de “Join” e um outro “Input Data” com as informações referentes ao “de-para” do código com a origem dos dados e a empresa responsável pela realização do transporte divulgado

pela *TLC*. O “Join” junta as duas bases e cria a nova coluna “Tipo” com o operador da corrida para cada um dos códigos fornecidos. Os registros que contêm os dados de operador formam a base que terá seu uso continuado e saem pela parte central da ferramenta. Pelo lado esquerdo, é gerada uma base com os registros sem operador, com o propósito de arquivamento. Este fluxo do Alteryx é repetido para todos os anos de 2015 a 2018 e sua versão da última data pode ser observada na figura 14.

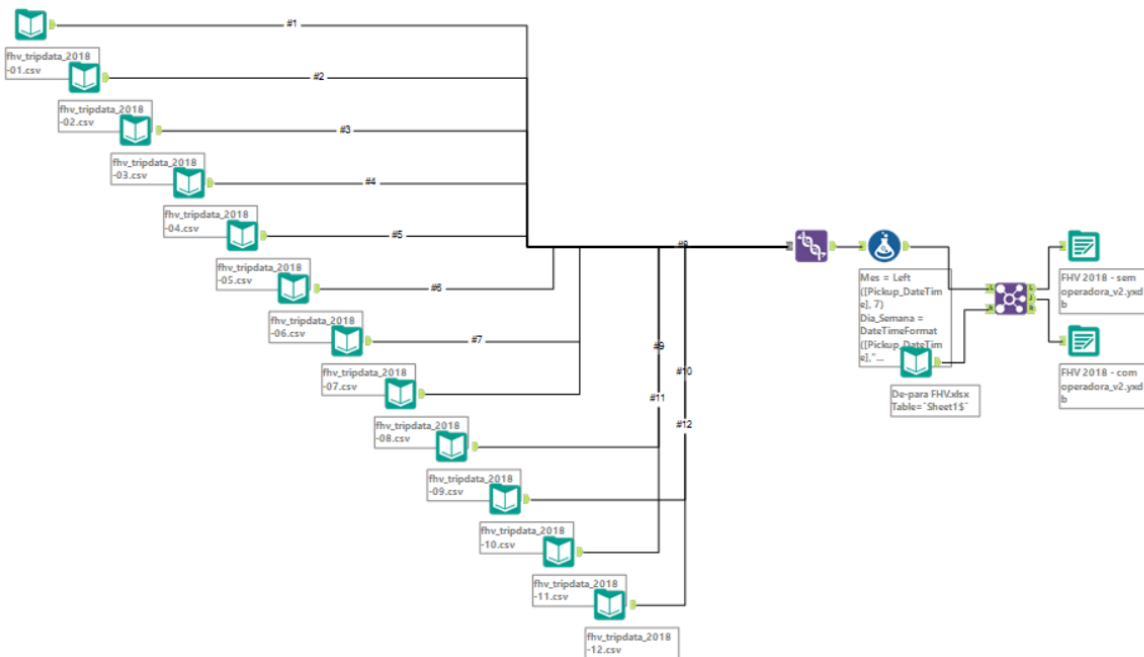


Figura 14: Fluxo do Alteryx que representa o tratamento inicial dos dados de *FHV* para o ano de 2018.
Fonte: elaborada pelo autor.

4.1.3. Dados de táxi com *pick-up* e *drop-off* por zona de táxi

Como mencionado na seção 3.2.1.1., de 2009 a 2016, as bases de dados de táxi são divulgadas com a latitude e a longitude ao invés das zonas de táxi correspondentes, tanto para o processo de *pick-up* quanto o de *drop-off*. Para tornar essas bases equiparáveis às de Uber e às bases de táxi mais recentes, é necessário cruzar essas informações com os *shapefiles .shp* da *TLC*. Para realizar esse ajuste, foram escolhidos os anos de 2009 e 2013, pelos motivos explicados em 4.1, e as bases já prontas desses anos, geradas em 4.1.1., foram submetidas ao fluxo da figura 15, exemplificado para o caso de 2009.

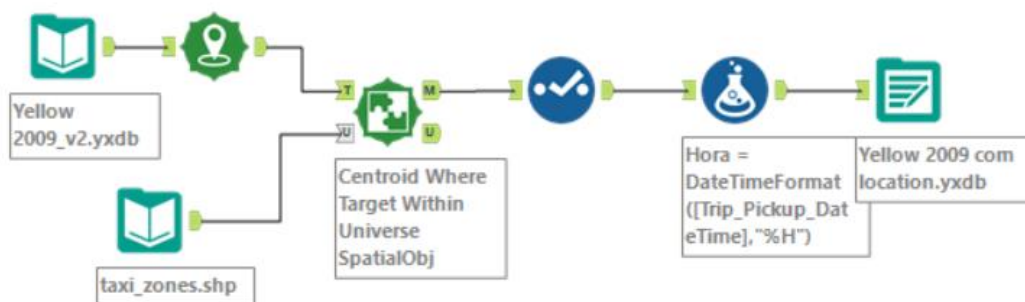


Figura 15: Fluxo do Alteryx que representa o tratamento adicional realizado nos dados de táxi para incluir as zonas de *pick-up* para o ano de 2009.

Fonte: elaborada pelo autor

No fluxo, a base pronta passa pela ferramenta de “Create Points”, que gera uma localização espacial para cada corrida lendo a latitude e a longitude do local de *pick-up*. Depois, através do comando “Spatial Match”, o software reconhece a qual área das zonas de táxi da *TLC* esse ponto pertence. Seguindo o fluxo, utiliza-se a ferramenta “Select” para retirar as informações espaciais, mantendo apenas o registro dos nomes das localidades de *pick-up*. Além disso, é gerada uma coluna apenas com o registro da hora em que houve o início dessa corrida, pois essa informação pode ser útil para avaliar a sazonalidade das corridas ao longo de um dia. Os registros espaciais foram retirados para diminuir o tamanho dos arquivos, que podia chegar a quase 500 gigabytes por ano somente com as informações de *pick-up*, mesmo no formato compactado de base de dados do Alteryx.

Depois de passar por esse fluxo, o arquivo passa novamente pelo mesmo fluxo para receber as informações das zonas espaciais do *drop-off*, sem criar uma nova coluna com o horário da corrida. As bases geradas de táxi e *FHV* de 2018 passam apenas por ajustes finos para serem compatíveis com as novas bases geradas. Nelas são adicionados os nomes das zonas de táxi e seus respectivos *boroughs* (nas bases originais tem apenas o código) e foi feita a fórmula para adicionar a coluna de “Hora”. No caso de *FHV*, os dados também foram filtrados para incluir apenas as corridas da Uber.

4.2. Análise descritiva dos dados

Antes de realizar análises mais avançadas, foi necessário entender um pouco mais dos dados e do panorama geral das corridas de táxi e *FHVs* em Nova Iorque.

No gráfico da figura 16, observa-se uma contagem simples do número de corridas mês a mês na cidade para táxis, Ubers, *FHVs* e o total de corridas (táxi somado com *FHVs*) na cidade. As partes em branco são os meses cujas informações da *TLC* estavam corrompidas,

como mencionado em 3.2.1.2. Pode-se observar que houve um período de estabilidade até o ano de 2014 e que, a partir deste, o número de corridas de táxi começa a cair gradualmente. Além disso, o mercado de Ubers e *FHVs* cresce em velocidade acelerada, assim como o total de corridas do mercado. Ademais, houve uma flutuação dos valores mês a mês, indicando que pode haver sazonalidade no número de corridas.

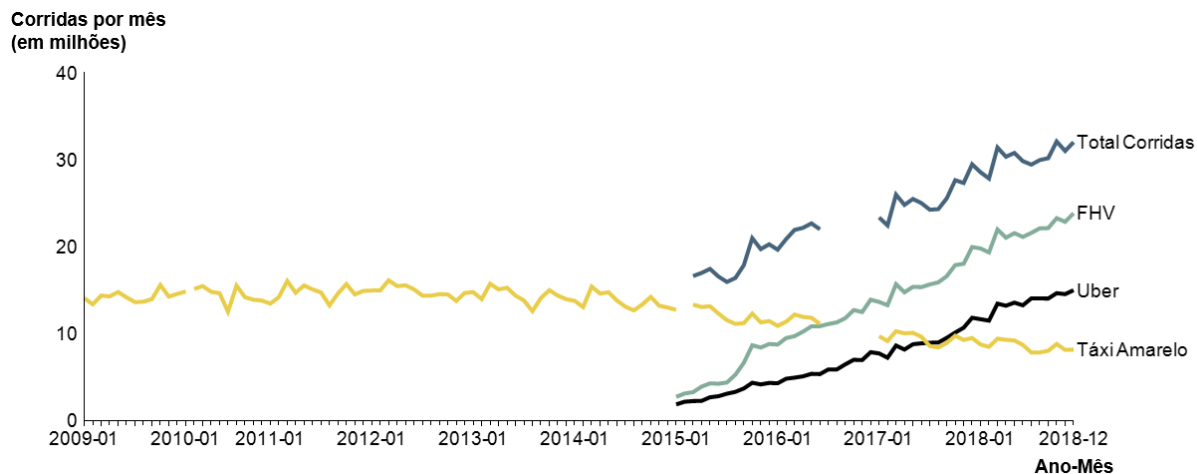


Figura 16: Número de corridas mês a mês realizadas em Nova Iorque por categoria de transporte individual.

Fonte: elaborada pelo autor.

Na figura 17, temos uma visão consolidada dos anos. De 2009 a 2013, o mercado ficou praticamente estável, apresentando um pouco de crescimento entre 2009 a 2012 e uma leve queda em 2013. Depois disso, houve um aumento relevante no número de corridas totais, que cresceu 110% entre 2013 e 2018. A queda em viagens de táxi iniciada em 2014 já indica a possível entrada de concorrentes no mercado, apesar da *TLC* ter iniciado a divulgar os dados de *FHVs* somente em 2015. Pode-se também observar que a Uber corresponde a maior parte das corridas de *FHVs* na cidade (163 milhões contra um total de 261). Para facilitar a visualização, nos anos em que não foi possível ler as informações de todos os meses de táxis, foi realizado um ajuste por meio de interpolação simples para chegar no total de corridas.

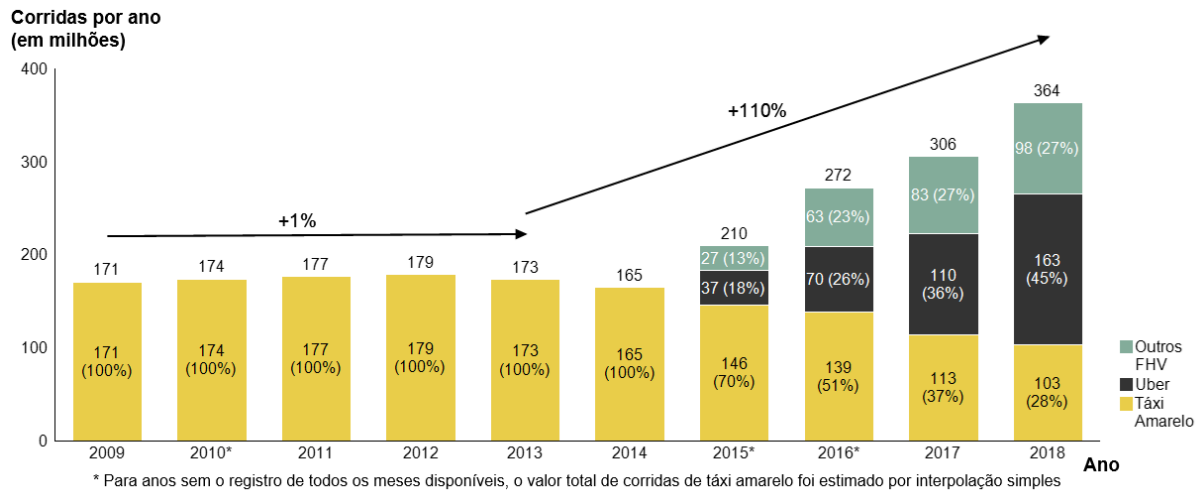


Figura 17: Número de corridas ano a ano realizadas em Nova Iorque por categoria de transporte individual.

Fonte: elaborada pelo autor.

Ao se observar somente os meses divulgados a partir de 2015, há indícios de que a queda no número de corridas de táxi está negativamente correlacionada com a entrada da Uber no mercado (figura 18). Com um cálculo simples de coeficiente de correlação de Pearson, utilizado de maneira linear para todos os meses de 2015 a 2018 com dados tanto para táxi quanto Uber, obteve-se uma correlação de -0,91. Entretanto, para se fazer afirmações mais assertivas seria necessário realizar um estudo econométrico mais aprofundado, de modo a considerar o efeito das sazonalidades e da dependência entre os meses.

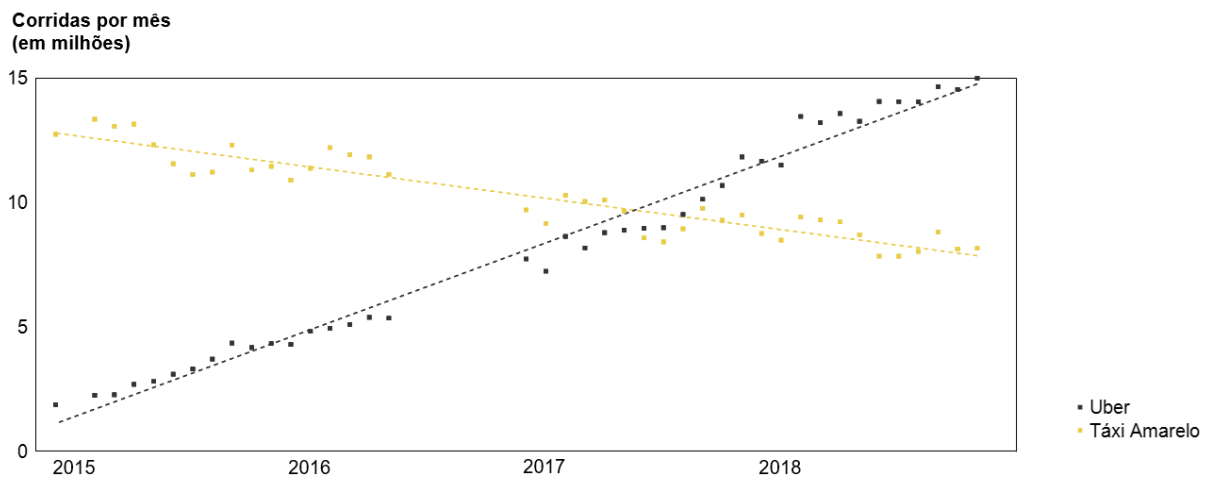


Figura 18: Comparação entre crescimento do número de corridas de Uber e queda das corridas de táxi em Nova Iorque entre 2015 e 2018.

Fonte: elaborada pelo autor.

4.3. Descrição do processo de análise

Para buscar diferenças no comportamento de táxi antes e depois da ascensão da Uber em Nova Iorque, foram realizadas 4 análises: (1) avaliação da variação nas distâncias e na duração das viagens; (2) investigação das mudanças na sazonalidade de corridas por dia da semana e por mês; (3) verificação de possíveis alterações nas zonas de táxi mais acessadas para *pick-up* e *drop-off* por dia da semana e por hora do dia; e (4) avaliação das diferenças entre origem e destino dos *boroughs* de Nova Iorque. Depois, realizou-se uma análise adicional para (5) medir a representatividade que os táxis mantiveram em Manhattan, por ser a área com maior concentração de corridas, para avaliar a adesão à Uber nesta região, independentemente dos padrões de deslocamento.

Essas análises primárias (1, 2 e 3) foram escolhidas baseando-se em avaliações vistas na revisão de literatura, que mostraram, por exemplo, que há sazonalidade nas corridas de táxi, e que a quantidade de viagens varia em função das diferentes zonas da cidade, como foi observado nas figuras 05, 06, 07, e 08 de Ferreira et al. (2013). As análises mais aprofundadas (04 e 05) foram realizadas em decorrência dos fatos aprendidos nas análises anteriores, que sugeriram investigar com mais profundidade os deslocamentos entre os *boroughs* e entender como se deu a penetração da Uber no mercado de corridas de Manhattan.

Em todas as análises, buscou-se sempre entender três fatores: (1) como o táxi era no início dos anos disponíveis e a variação de comportamento dentro deles, para avaliar a situação inicial, e se já havia alguma tendência de mudança antes da queda no número de corridas devida à entrada da Uber (entre 2009 a 2013); como (2) o táxi se tornou no final do período com dados disponíveis, para avaliar se houve alguma mudança depois da entrada do concorrente privado; e qual a (3) característica do comportamento das viagens de Uber, para avaliar se houve alguma aproximação ou distanciamento das características das corridas de táxi em caso de mudança.

4.3.1. Distância e duração das viagens

Para calcular a distância e a duração das viagens, foi gerada uma base de dados com a mediana da distância e da duração das corridas de táxi mês a mês de 2009 a 2018, com restrição apenas dos meses que não estão disponíveis nas bases. Para a Uber, foi calculada somente a duração das viagens para um número mais restrito de meses dado que não há informações de distância e que o *timestamp* de *drop-off* só passou a ser divulgado em julho de 2017. Foi utilizada a mediana para evitar distorções de linhas de dados com valores muito altos, como mencionado em 3.2.1.2.

No fluxo da figura 19, o software acessa as bases de táxi com os dados das corridas de cada ano, cria uma coluna de “Duração”, calculando a diferença de tempo em segundos entre o *drop-off* e o *pick-up* de cada corrida. Depois, os nomes das colunas com as informações de distâncias das corridas nativas da base são renomeados através do comando “Select” para que sejam compatíveis entre todos os anos, dado que há uma variação entre os nomes das colunas divulgadas ano a ano. Em seguida, a ferramenta de “Summarize” calcula a mediana dos valores de duração e distância das corridas para cada mês do ano; é criada uma coluna com o ano dos números para facilitar a futura identificação; e são filtrados os meses somente do ano vigente de cada base para remover possíveis *outliers*, como explicado também na seção 3.2.1.2. Em seguida, o “Union” une todas as informações em uma mesma base, já pronta para a construção dos gráficos com os resultados utilizando o Excel e o Mekko Graphics.

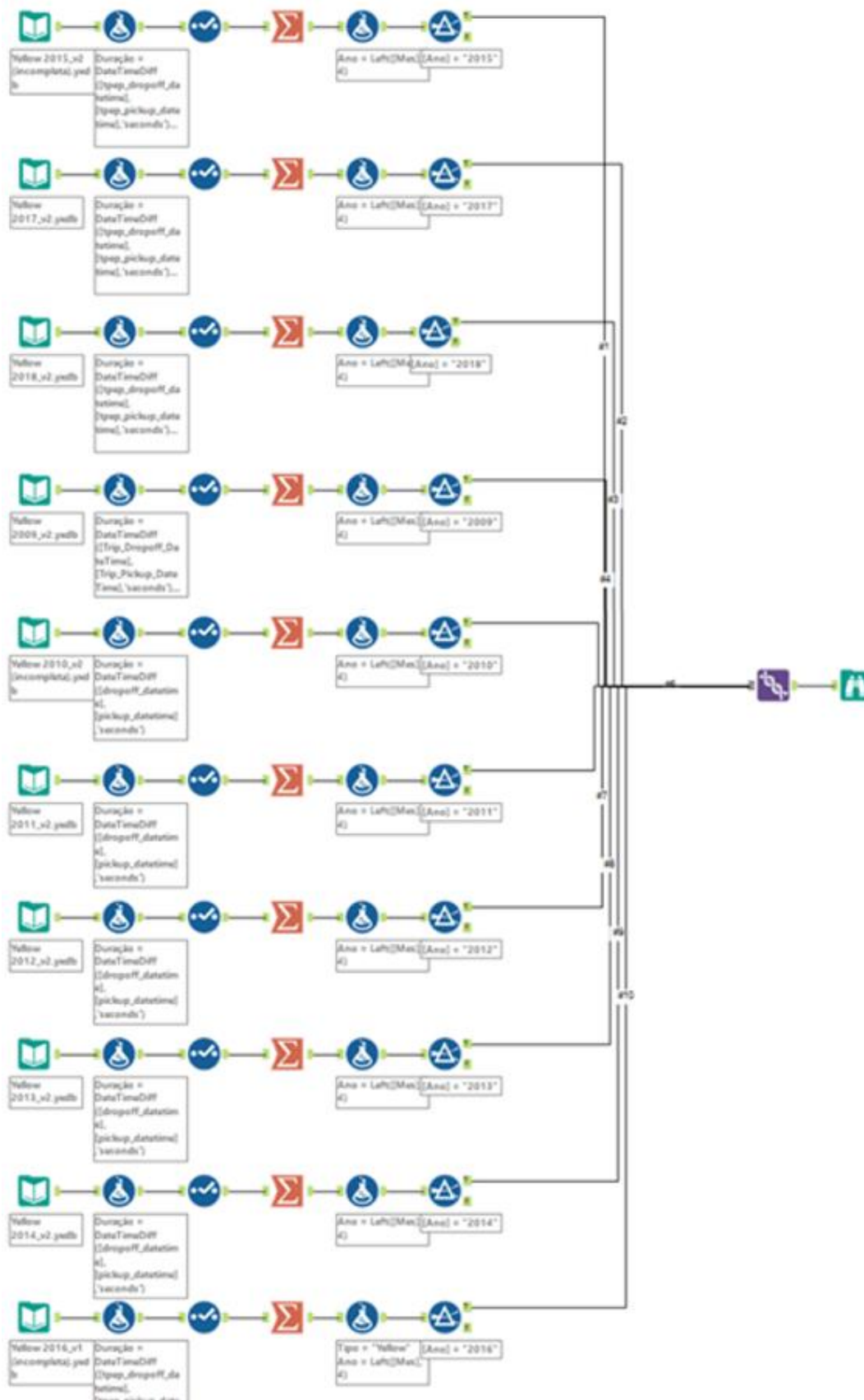


Figura 19: Fluxo do Alteryx utilizado para calcular a distância e a duração das viagens de táxi.
Fonte: elaborada pelo autor.

Já a figura 20 representa o fluxo para o cálculo da duração das corridas de Uber. Primeiro são filtrados somente os dados dessa empresa na base de *FHVs* identificados. Depois foram calculadas as durações das corridas e as medianas por mês. Especificamente em 2017, foi colocado um filtro para remover as corridas que ainda não tem o *timestamp* de *drop-off* divulgado, sendo elas de janeiro a junho do respectivo ano.

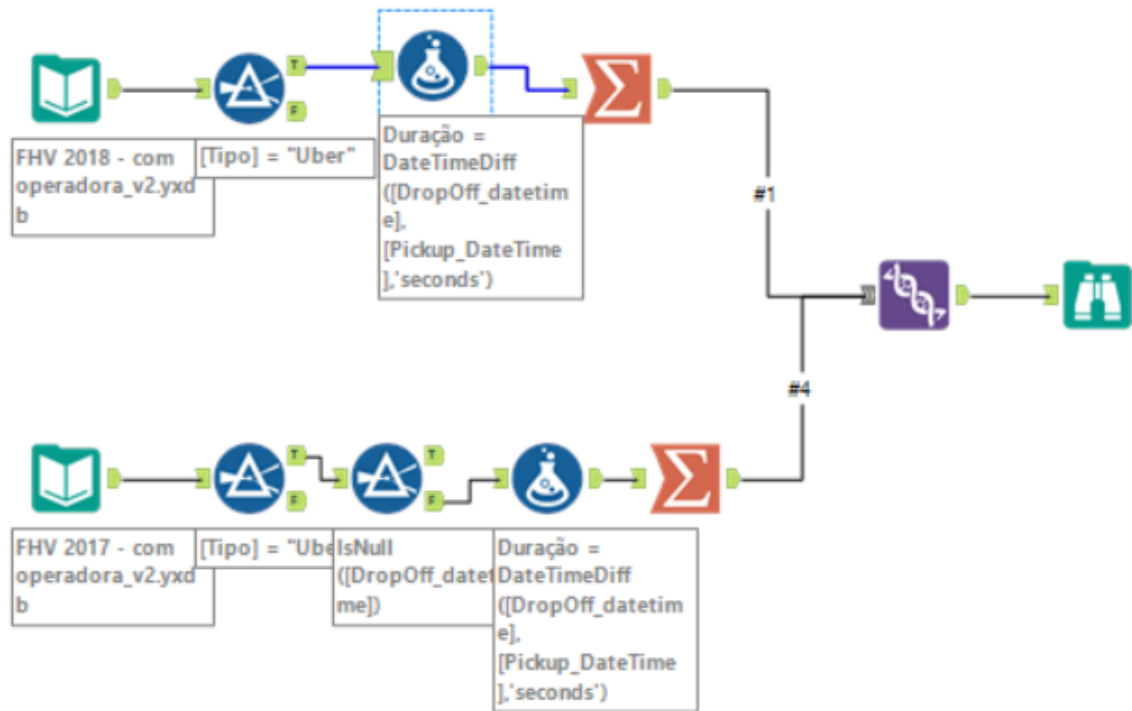


Figura 20: Fluxo do Alteryx utilizado para calcular a distância e a duração das viagens de Uber.

Fonte: elaborada pelo autor.

4.3.2. Sazonalidade por dia da semana e mês

Para analisar a sazonalidade por dia da semana e mês, foi gerada no Alteryx uma base com o número de corridas por mês e por dia da semana para todos os meses disponíveis de táxi e Uber. Depois, foram realizadas análises simples no Excel e os dados plotados pelo Mekko Graphics. No caso da análise para dias da semana de táxi, após realizados os cálculos, os anos sem todos os meses disponíveis foram ajustados por meio de uma regressão linear para que o total de corridas fosse mais coerente com a ordem de grandeza da quantidade de viagens que ocorreram, de modo a facilitar a visualização.

Para o táxi, foi utilizado o fluxo da figura 21. Nele, a base de cada ano é lida, depois é criada uma coluna com o ano do *pick-up* de cada registro para facilitar as análises. A seguir, o ano é filtrado, de modo a atender a remoção de *outliers* proposta na seção 3.2.1.2. Depois, é

realizada uma contagem do número de corridas, agrupado por dia da semana e por mês. Por último, os dados são unidos em uma única base, contendo as contagens de todos os anos.

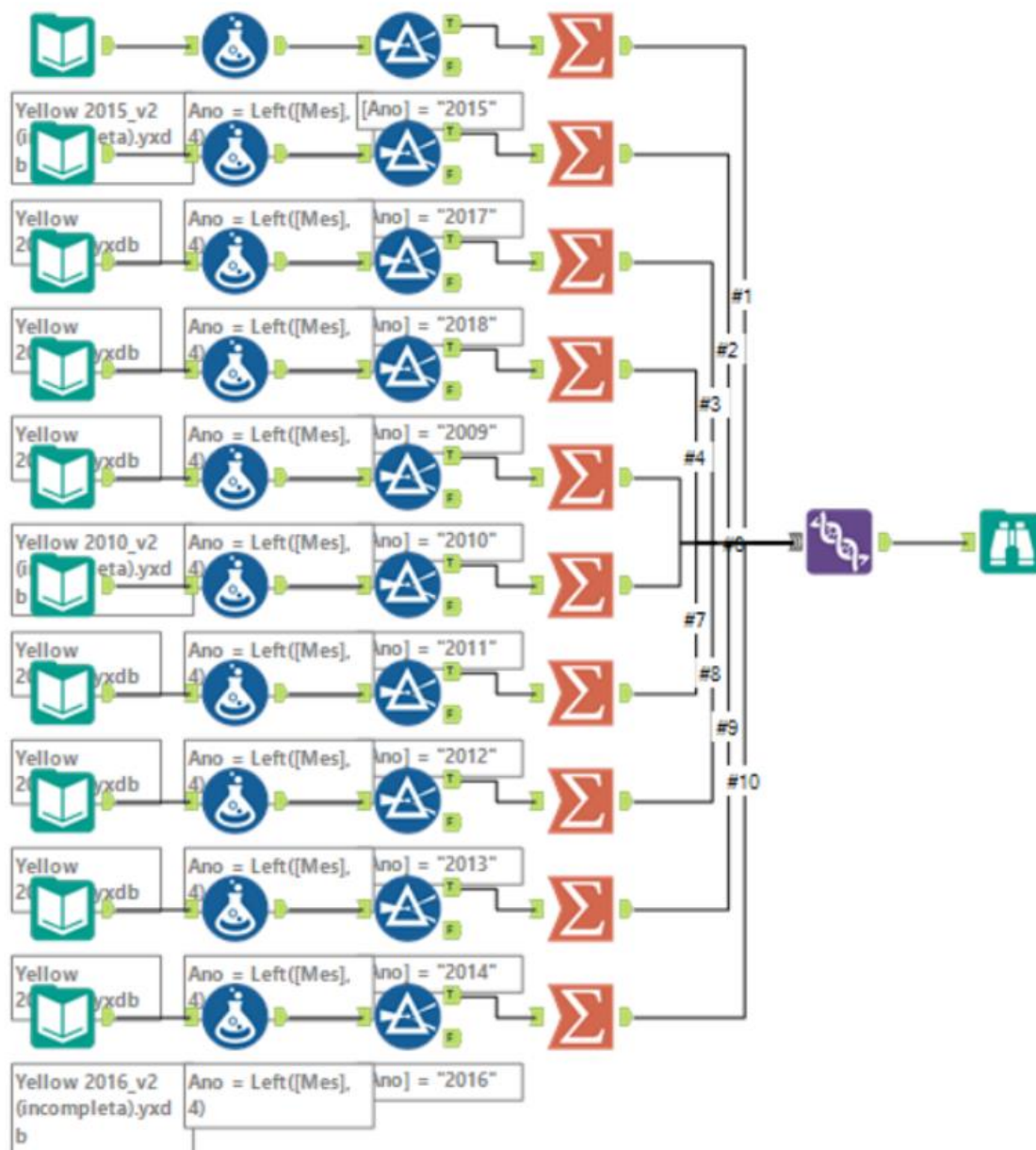


Figura 21: Fluxo do Alteryx utilizado para calcular a sazonalidade por dia da semana e mês das viagens de táxi.

Fonte: elaborada pelo autor.

Para os dados de Uber, é aberta cada base com as operadoras e filtrados os dados correspondentes à Uber. Depois, é feito o mesmo agrupamento dos táxis e junção das bases (figura 22). Nesse caso, não foi necessário remover os *outliers* pois o mesmo erro dos táxis não ocorre na base de *FHV*.

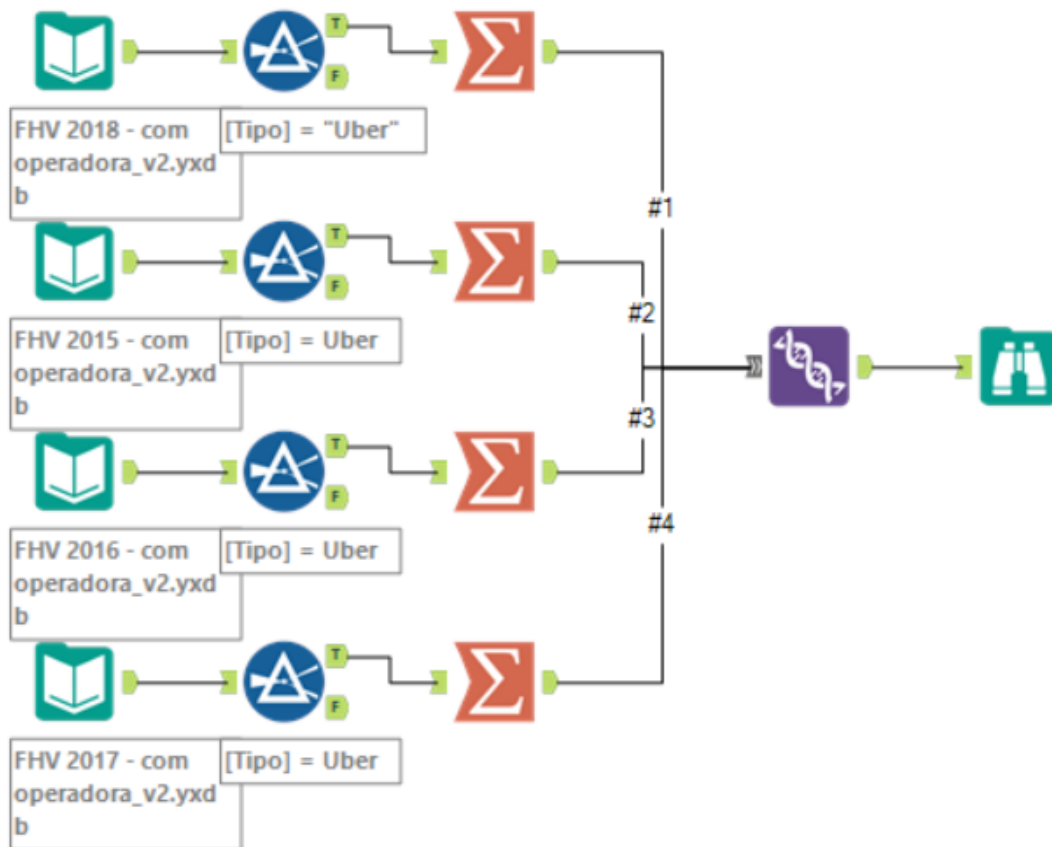


Figura 22: Fluxo do Alteryx utilizado para calcular a sazonalidade por dia da semana e mês das viagens de Uber.

Fonte: elaborada pelo autor.

4.3.3. Número de viagens por zona de táxi, dia da semana e horário

Para essa análise, foram utilizadas as bases geradas na seção 4.1.3., que possuem as zonas de táxi de *pick-up* e *drop-off* para todos os anos. Em todos os casos, foi adicionada uma coluna “Ano” e foram sumarizadas as bases por número de viagens por dia da semana, hora do dia do *pick-up* e zona de táxi, como pode ser observado na figura 23. Esse procedimento foi feito para gerar uma base com as informações de *pick-up* e depois para gerar outra base com as informações de *drop-off*. Dado que as duas informações já existem nos *inputs*, poderia ter sido gerada apenas uma base com as duas informações, mas a base resultante teria um número de linhas muito grande, logo optou-se por quebrar em duas bases. Em ambos os dois casos, a base resultante contou tanto com as informações tanto de táxi quanto com as de Uber.

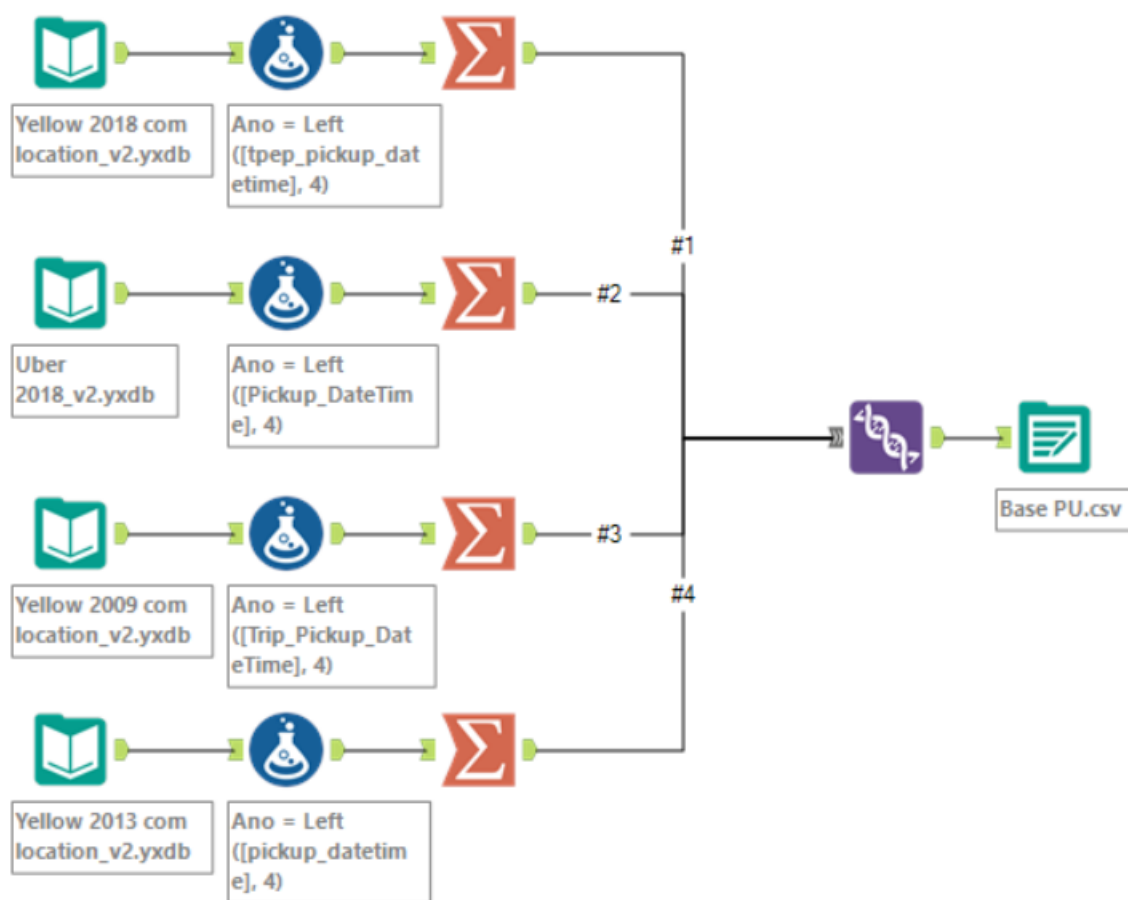


Figura 23: Fluxo do Alteryx utilizado para visualizar as viagens por zona, dia da semana e horário das viagens de táxi e Uber.

Fonte: elaborada pelo autor.

Para visualizar as informações, foi utilizado o software Tableau. Nele, como as informações espaciais foram perdidas na elaboração da seção 4.1.3., ficando somente as informações referentes a identificação das zonas de táxi, foi necessário cruzar novamente essa informação com as suas áreas espaciais, por meio do arquivo *.shp* da *TLC*, como pode ser visto na figura 24.

Tableau - Book1

FileDataServerWindowHelp

Connections

Base PU
Text file

taxi_zones
Spatial file

Files

taxi_zones.shp

Base PU.csv+ (Multiple Connections)

Connection
LiveExtract

Filters
0Add

Base PU.csv

taxi_zones.shp

Sort fields

Data source order

Show aliasesShow hidden fields1,000rows

Base PU.csv	Base PU.csv	Base PU.csv	Base PU.csv	Base PU.csv	Base PU.csv	Base PU.csv	Base PU.csv	taxi_zones.shp	taxi_zones.shp	taxi_zones.shp	taxi_zones.shp	taxi_zones.shp	taxi_zones.shp
Tipo	Ano	Dia Semana	Hora	PU Borough	PU Zone	Pu Id	Count PU	Objectid	Shape Leng	Shape Area	Zone	Location ID	Bo
Uber	2018	Thursday		1	Manhattan	Midtown Center	161	9.071	161	0.035804	0.00007191	Midtown Center	161
Uber	2018	Thursday		1	Manhattan	Midtown East	162	4.173	162	0.035270	0.00004789	Midtown East	162
Uber	2018	Thursday		1	Manhattan	Midtown North	163	3.595	163	0.034177	0.00004086	Midtown North	163
Uber	2018	Thursday		1	Manhattan	Midtown South	164	6.421	164	0.035772	0.00005563	Midtown South	164
Uber	2018	Thursday		1	Manhattan	Morningside Hei...	166	1.932	166	0.068824	0.00025528	Morningside Hei...	166
Uber	2018	Thursday		1	Manhattan	Murray Hill	170	3.421	170	0.045769	0.00007432	Murray Hill	170
Uber	2018	Thursday		1	Manhattan	Penn Station/Ma...	186	2.204	186	0.024696	0.00003707	Penn Station/Ma...	186
Uber	2018	Thursday		1	Manhattan	Randalls Island	194	5	194	0.081794	0.00027348	Randalls Island	194
Uber	2018	Thursday		1	Manhattan	Roosevelt Island	202	166	202	0.069922	0.00010618	Roosevelt Island	202
Uber	2018	Thursday		1	Manhattan	Seaport	209	1.059	209	0.030695	0.00003887	Seaport	209

Go to Worksheet

Data Source

Sheet1

2015

05/08/2019

Figura 24: Conexão da base gerada com as viagens por zona, dia da semana e horário das viagens de táxi e Uber com a base de zona de táxi definidas pela TLC.

Fonte: elaborada pelo autor.

4.3.4. Localização de *pick-up* e *drop-off*

Para a análise dos fluxos de deslocamento, foram utilizadas novamente as bases da seção 4.1.3. Elas, como mostra a figura 25, são lidas, sumarizadas pelo agrupamento de *borough* de *pick-up* e de *drop-off*, e foi gerada a contagem de corridas realizadas em cada combinação origem-destino. Em seguida, é marcado o ano e o tipo de cada base, e elas são unidas.

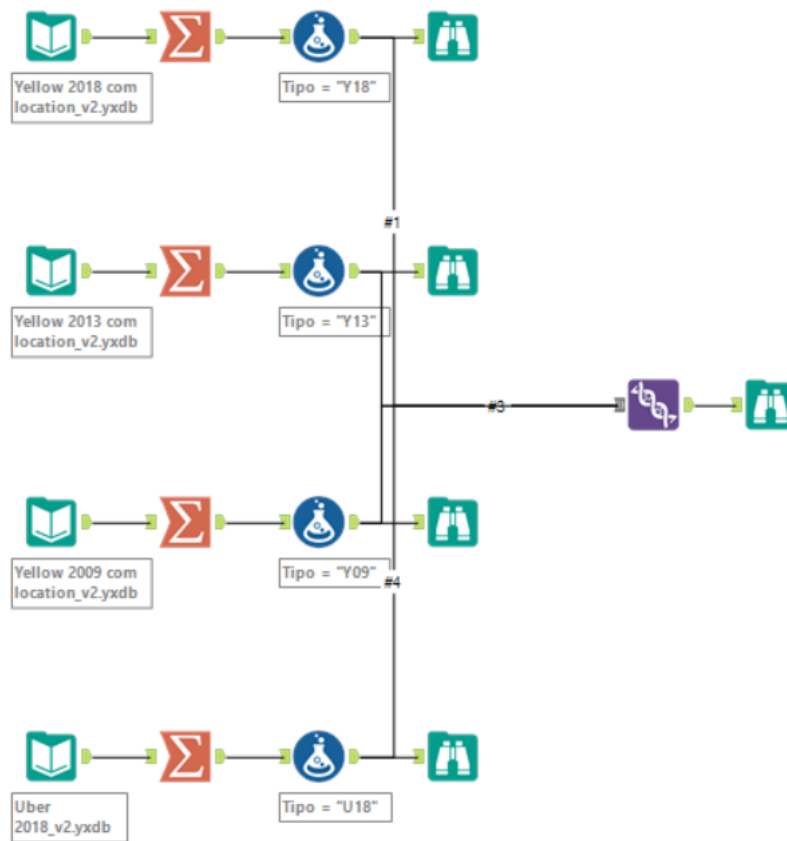


Figura 25: Fluxo do Alteryx utilizado para analisar a recorrência de origem-destino das viagens de táxi e Uber.

Fonte: elaborada pelo autor.

A base resultante foi analisada por meio da elaboração de diagramas de Sankey no Tableau.

4.3.5. Evolução das corridas em Manhattan

Para avaliar a evolução das corridas que acontecem no *borough* de Manhattan, foram utilizadas as bases da seção 4.1.3. Como estas não foram geradas para *FHVs*, a análise neste caso irá partir das bases originais geradas no 4.1.2.

Primeiro, foram feitas bases com todas as corridas que contêm o *pick-up* ou o *drop-off* em Manhattan, como pode ser observado na figura 26. O fluxo de Alteryx cria uma coluna de ano e filtra os valores que não correspondem ao ano original para remover os outliers. Depois, são aplicados dois filtros. No primeiro, separam-se todas as corridas com o *pick-up* em Manhattan. Em seguida, para as que se iniciaram em outro *borough*, são filtradas aquelas que tiveram o *drop-off* em Manhattan. Dessa forma, captura-se todas as corridas que tiveram o *pick-up* ou o *drop-off* (ou os dois) no bairro de interesse. Depois, são unidas todas as bases e é feita uma contagem para o número de corridas verificadas em cada ano.

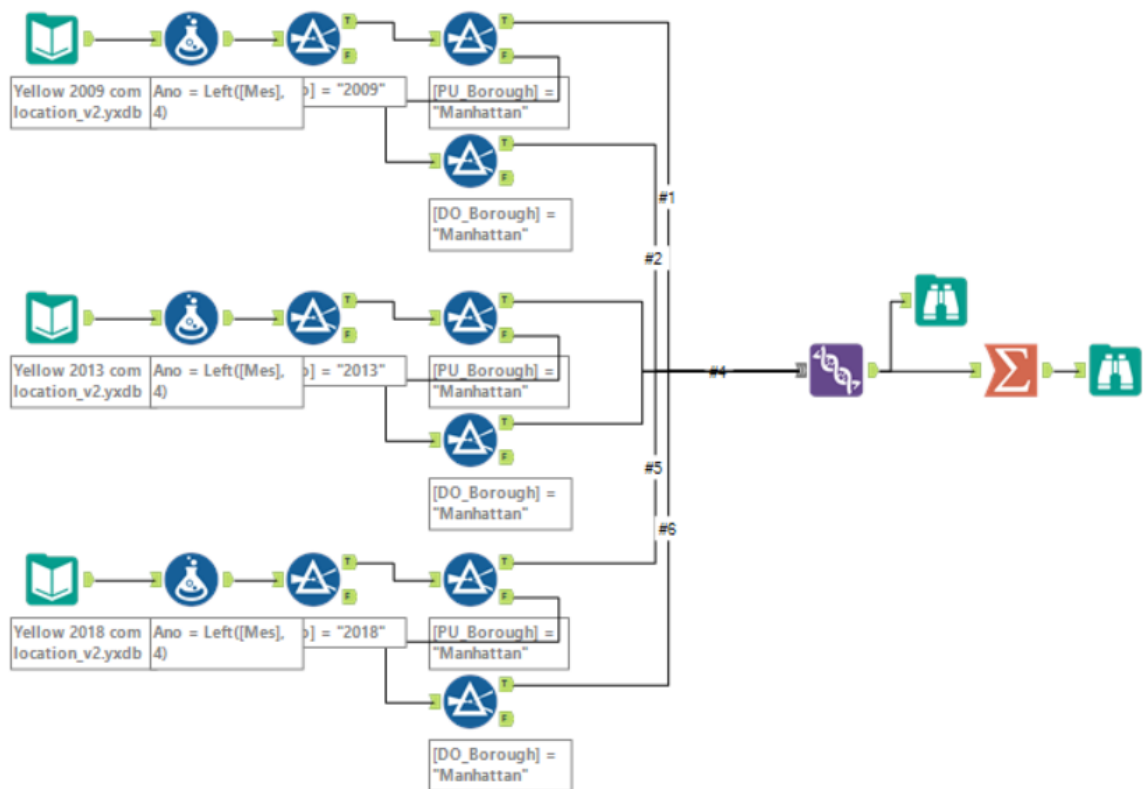


Figura 26: Fluxo do Alteryx utilizado para analisar a evolução das corridas de táxi que começaram ou terminaram em Manhattan.
Fonte: elaborada pelo autor.

A segunda base, como observado na figura 27, foi feita por um processo parecido com o da figura anterior. Entretanto, ao invés de capturar todas as corridas que passam por Manhattan, filtram-se as corridas com o *drop-off* dentro desta região no conjunto de dados das viagens que tiveram o *pick-up* nesta mesma localidade. Dessa forma, obteve-se uma base com a contagem das corridas que começaram e terminaram no *borough*, focando ainda mais no local de interesse.

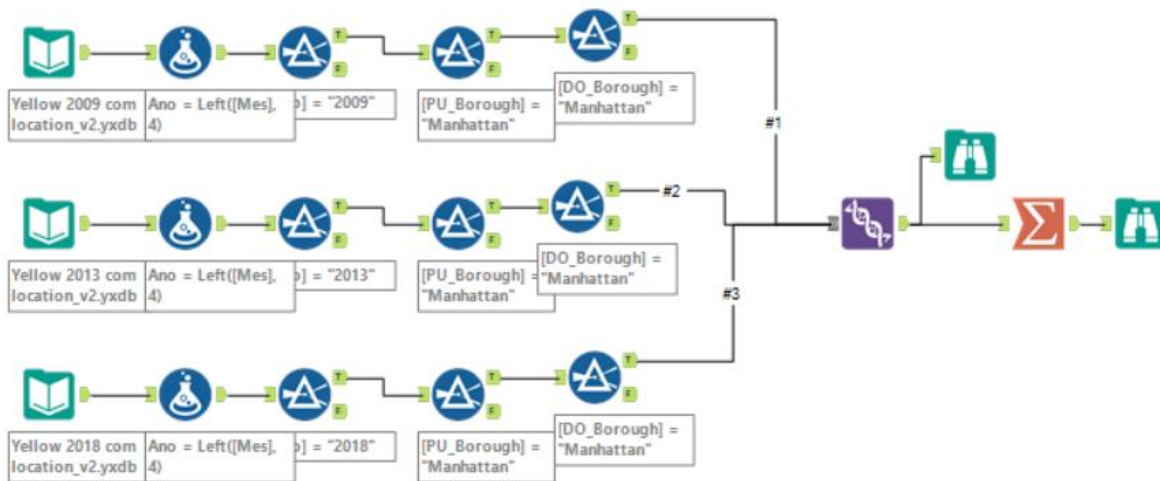


Figura 27: Fluxo do Alteryx utilizado para analisar a evolução das corridas de táxi que começaram e terminaram em Manhattan.

Fonte: elaborada pelo autor.

Já as corridas de *FHV* passam por um processo semelhante ao das bases da seção 4.1.3., no qual elas primeiro recebem os nomes das zonas de táxi (que já estão marcadas com uma identificação na base original). Para não deixar de capturar registros em que apenas o *pick-up* ou o *drop-off* estão marcados na base, une-se as informações das corridas que não tiveram *match* com o “de-para” da base de zonas de táxi da *TLC*. Nesse processo, há uma perda de 11% dos registros de *FHV* que acabam não sendo considerados nas análises por falta de informação. Depois disso, o fluxo cria duas rotas diferentes, uma para cada um dos filtros das figuras 26 e 27, gerando uma base com os dados de *FHV* que começam e/ou terminam em Manhattan e outro com os dados que começam e terminam no *borough*, quebrados pela operadora do transporte. O processo pode ser observado na figura 28.

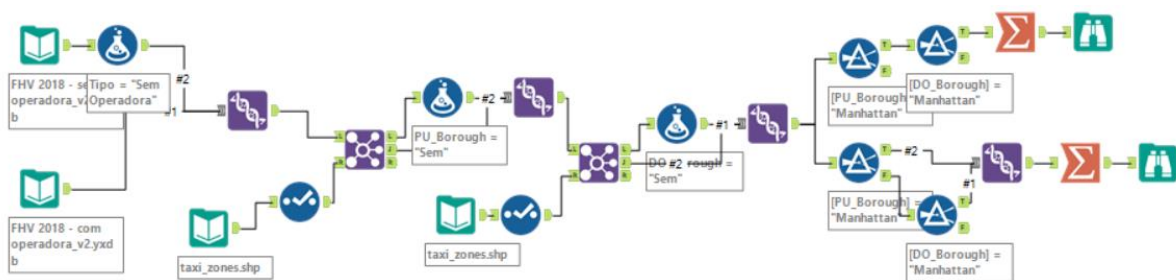


Figura 28: Fluxo do Alteryx utilizado para analisar a evolução das corridas de Uber que começaram e/ou terminaram em Manhattan e as corridas que começaram e terminaram neste *borough*.

Fonte: elaborada pelo autor.

4.4. Resultado das análises

4.4.1. Distância e duração das viagens

Na figura 29, podemos visualizar a variação da mediana das distâncias percorridas pelos táxis mês a mês de 2009 a 2018. As partes em branco são os dados que não estão disponíveis para análise e não há informações de Uber para distância no período.

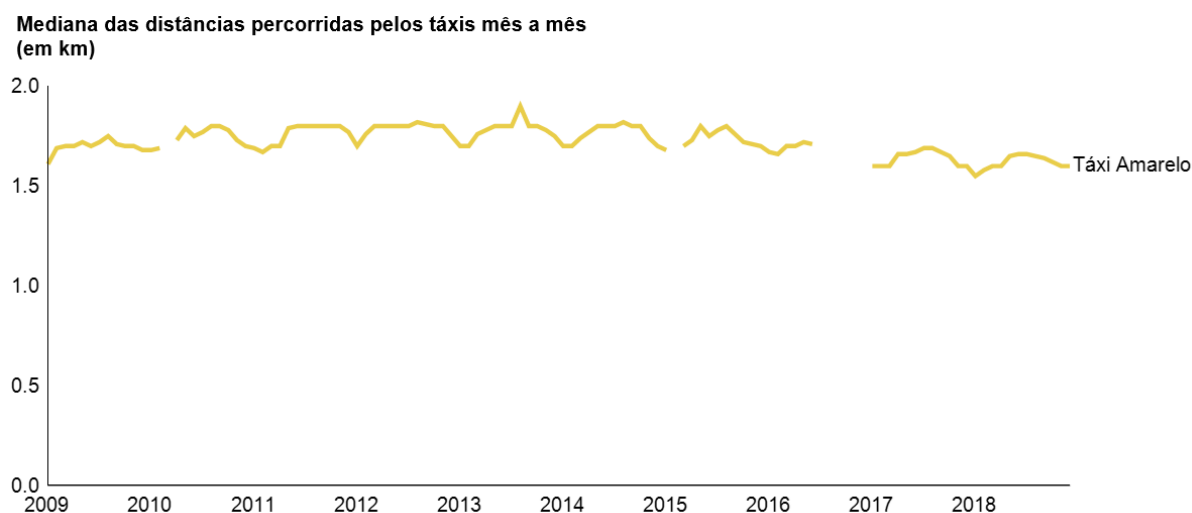


Figura 29: Mediana das distâncias percorridas pelos táxis mês a mês.
Fonte: elaborada pelo autor com dados da *TLC*.

Ao verificar os dados, percebe-se que a mediana é relativamente estável ao longo dos anos, apresentando um leve crescimento durante 2009, um pico em agosto de 2013 e uma tendência de queda a partir da metade de 2015, passando de uma mediana média de 1,74 quilômetros em 2015 para 1,62 em 2018.

No gráfico da figura 30, tem-se a mediana da duração das viagens de táxi para o mesmo período e das corridas de Uber de julho de 2017 até o final de 2018.

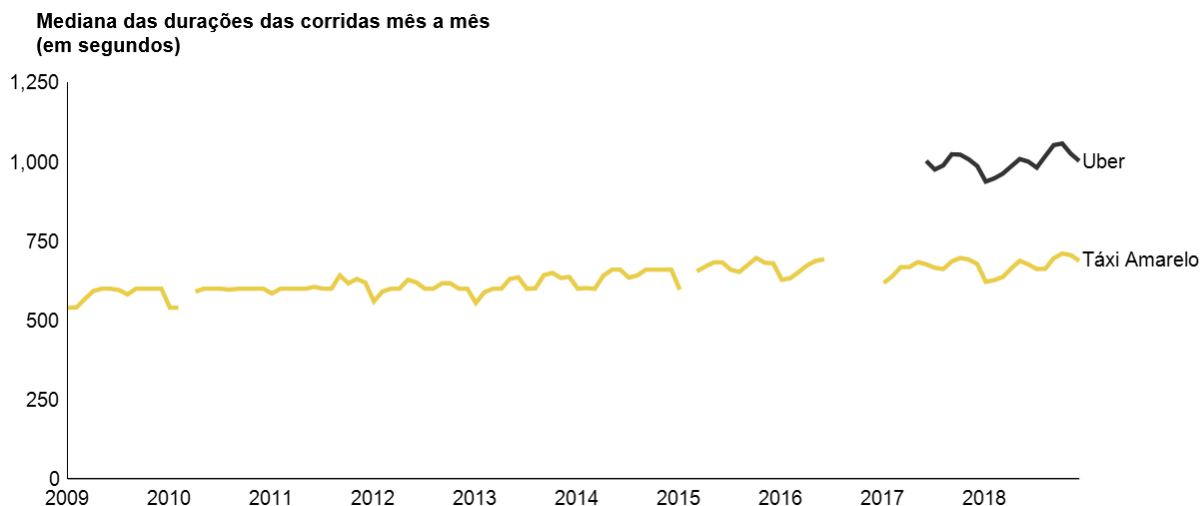


Figura 30: Mediana das durações das corridas de táxi e Uber mês a mês.

Fonte: elaborada pelo autor com dados da *TLC*.

Enquanto houve uma queda da mediana das distâncias, a mediana das durações das viagens aumenta em ritmo desacelerado, porém frequente, ao longo da série temporal, passando de uma mediana média de 585 segundos em 2009 para 670 em 2018. Apesar de haver uma tendência de queda na distância percorrida pelas viagens, há uma tendência de aumento na duração das corridas. Além disso, as corridas de Uber são consideravelmente mais longas em duração que as de táxi (de 45% a 54% mais duradouras), apresentando sazonalidade mês a mês visualmente semelhante (variações para cima e para baixo acompanham os mesmos meses, como observa-se no gráfico).

Apesar de haver variações ano a ano nos índices calculados para os táxis, não se encontrou nenhuma evidência de mudança de distância e duração das viagens devido a entrada da Uber no mercado.

4.4.2. Sazonalidade por dia da semana e mês

4.4.2.1. Sazonalidade por dia da semana

Na figura 31, visualiza-se a quantidade total de corridas de Uber por ano realizadas por dia da semana. Observa-se um aumento da quantidade de corridas ao longo dos dias úteis da semana, de segunda a sexta, em todos os anos. Depois, há um pico de demanda no sábado e uma caída nas corridas no domingo.

Corridas de Uber por dia da semana por ano
(em milhões)

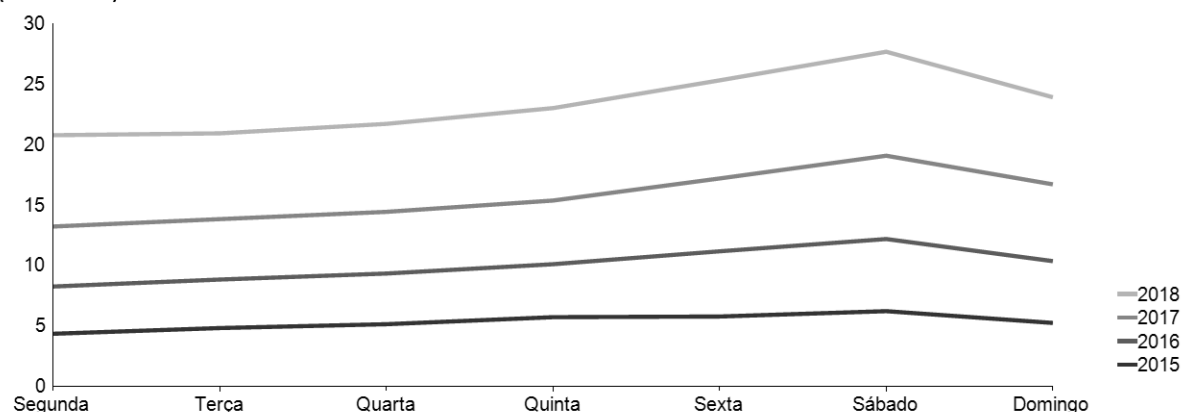
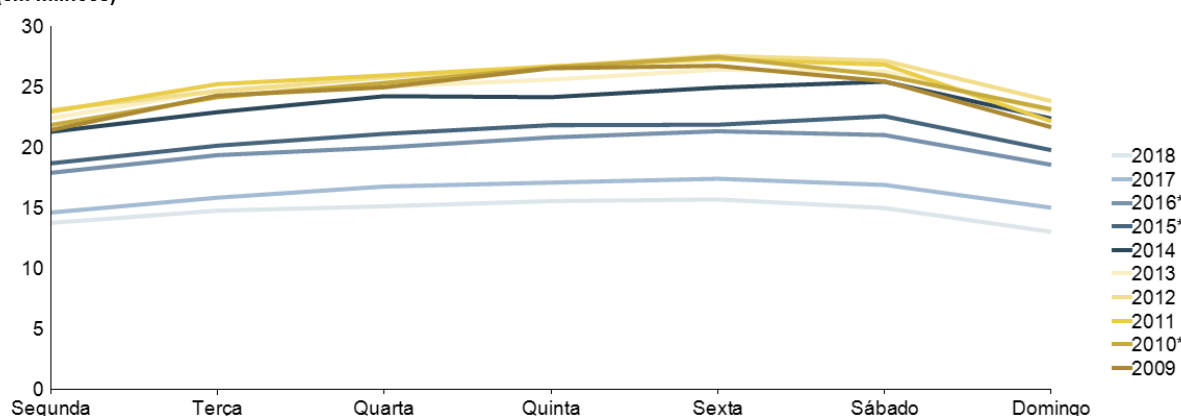


Figura 31: Corridas de Uber por dia da semana e por ano.

Fonte: elaborada pelo autor com dados da TLC.

Na figura 32, observa-se os mesmos dados para táxi. Os números seguem a mesma tendência da Uber, com um crescimento no número de corridas até sábado e uma queda no domingo, apesar de apresentarem uma sazonalidade menor, com uma variação menos intensa aos sábados (média da variação entre o número de corridas em cada dia da semana e a média dos 7 dias para todos os anos foi de 9,7% para a Uber e 5,7% para os táxis e variação máxima foi de 21,6% contra 12,4%). Os dados dos primeiros anos da série estão bastante sobrepostos, devido ao número total de corridas semelhante. Conforme cai a quantidade de corridas pela diminuição da demanda do táxi, as curvas vão ficando mais espaçadas, mas seguem apresentando a mesma tendência.

Corridas de táxi amarelo por dia da semana por ano
(em milhões)



* Para anos sem o registro de todos os meses disponíveis, o valor total de corridas de táxi amarelo foi estimado por interpolação simples

Figura 32: Corridas de táxi por dia da semana e por ano.

Fonte: elaborada pelo autor com dados da TLC.

Não há indícios de mudança na sazonalidade semanal dos táxis devido a entrada da Uber na cidade.

4.4.2.1. Sazonalidade por mês

Na figura 33, verifica-se o número de corridas de Uber por mês ao longo dos anos. Nesse caso, como existe uma tendência de crescimento dos números ao longo de um ano, não se tirou conclusões a respeito da sazonalidade mensal.

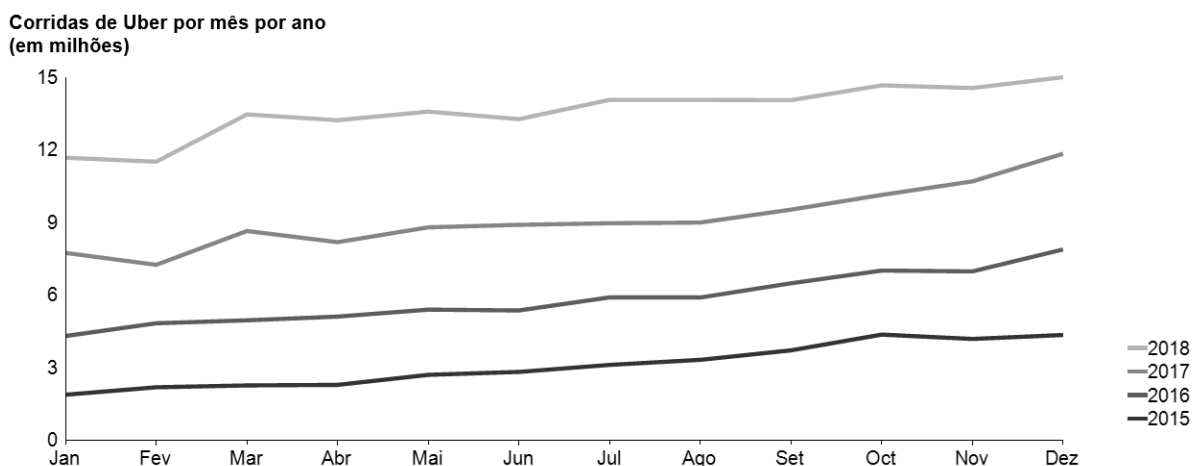


Figura 33: Corridas de Uber por mês e por ano.
Fonte: elaborada pelo autor com dados da *TLC*.

Para as informações de táxi, dispostas na figura 34, há uma oscilação para cima nos meses de fevereiro, março e abril, e uma oscilação para baixo nos meses de julho, agosto e setembro. Essa oscilação apresenta variações entre os anos, mas todos eles têm uma curva geral com essas características. A partir do ano de 2015, há uma tendência de queda do número total de corridas acentuada devido a entrada da concorrente e curvas dos anos vão se distanciando da sobreposição dos anos iniciais. Mesmo assim, elas mantêm uma curva semelhante à dos anos anteriores, de modo que não se pode perceber nenhuma mudança de sazonalidade nos meses do ano devida à ascensão do concorrente privado.

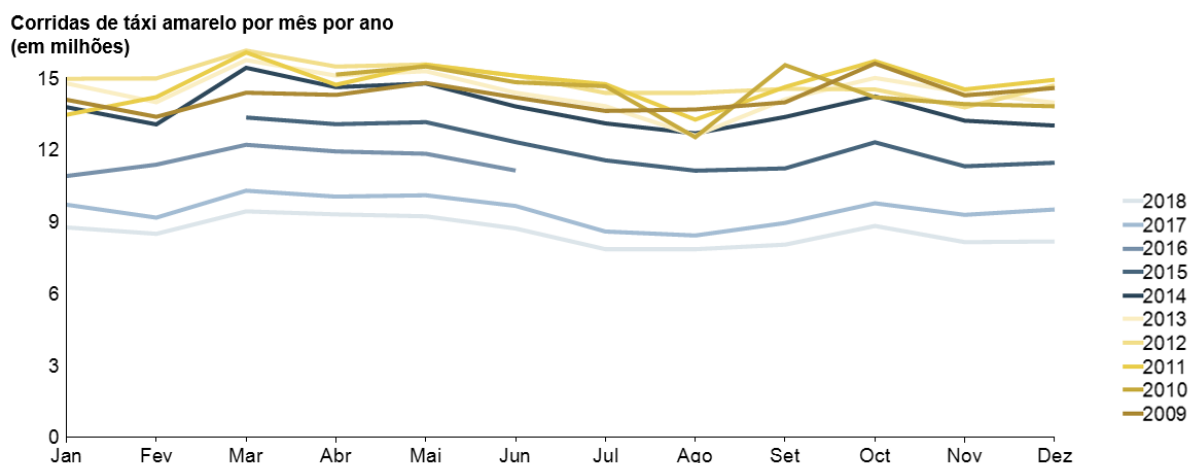


Figura 34: Corridas de táxi por mês e por ano.
Fonte: elaborada pelo autor com dados da TLC.

4.4.3. Número de viagens por zona de táxi, dia da semana e horário

Para a análise do número de viagens por zona de táxi por dia da semana de zona de táxi por horário, foram feitos mapas de calor com a quantidade de viagens realizadas em cada zona de táxi nos anos de 2009, 2013 e 2018 e para Uber em 2018. Há versões do mapa tanto para o *pick-up* quanto para o *drop-off* de passageiros, e para a variação de acordo com as horas do dia e com os dias da semana.

Devido a restrições para disponibilizar a animação desses dados em papel, eles estão disponíveis nos *links* correspondentes na tabela 3.

Tabela 03: *Links* para visualização da análise do número de viagens por zona de táxi, dia da semana e horário.

Localização visualizada	Categoria	Site
<i>Pick-up</i>	Hora do Dia	https://public.tableau.com/profile/bruno4221#!/vizhome/Pick-up-Workbook/Hora
<i>Pick-up</i>	Dia da Semana	https://public.tableau.com/profile/bruno4221#!/vizhome/Pick-up-Workbook/Dia
<i>Drop-off</i>	Hora do Dia	https://public.tableau.com/profile/bruno4221#!/vizhome/Drop-off-Workbook/Hora
<i>Drop-off</i>	Dia da Semana	https://public.tableau.com/profile/bruno4221#!/vizhome/Drop-off-Workbook/Dia

Fonte: elaborado pelo autor.

Para acessar os dados, basta digitar o *link* no navegador de sua preferência, clicar na opção de visualização em tela cheia, como representado na figura 35, e depois variar a métrica desejada conforme observado na figura 36.

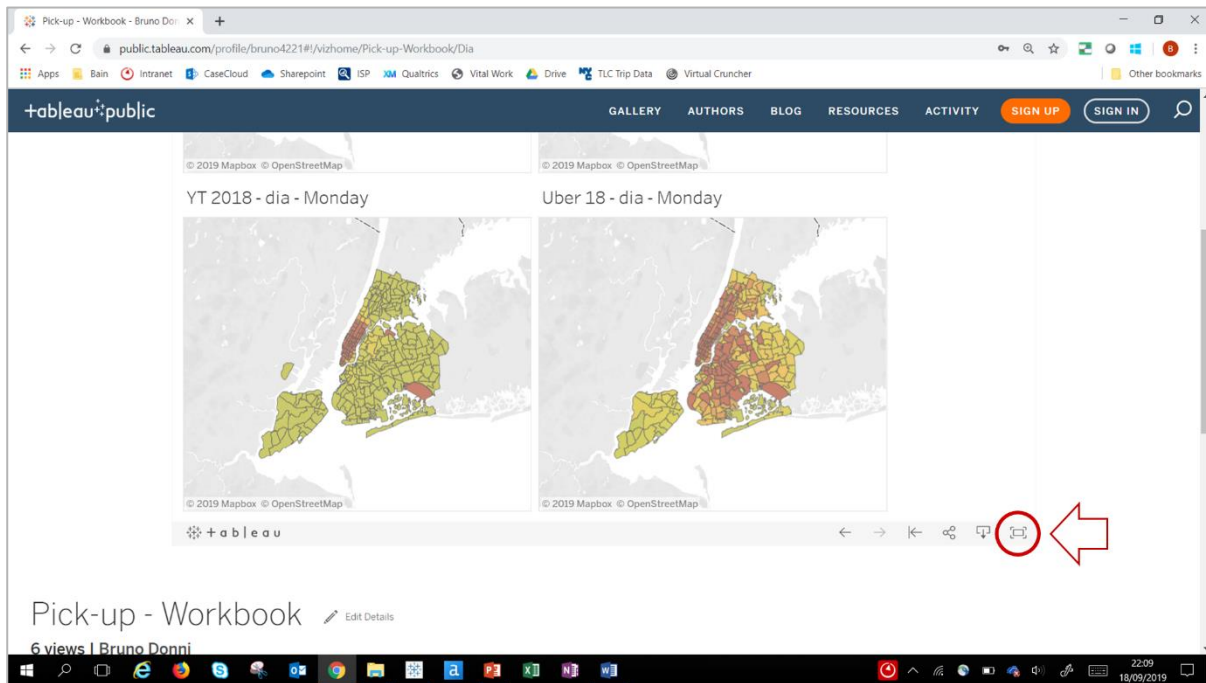


Figura 35: Indicação do botão utilizado para visualizar os mapas em tela cheia dentro da página do Tableau Public com a análise do número de *pick-ups* de táxi por dia da semana.
Fonte: elaborada pelo autor.

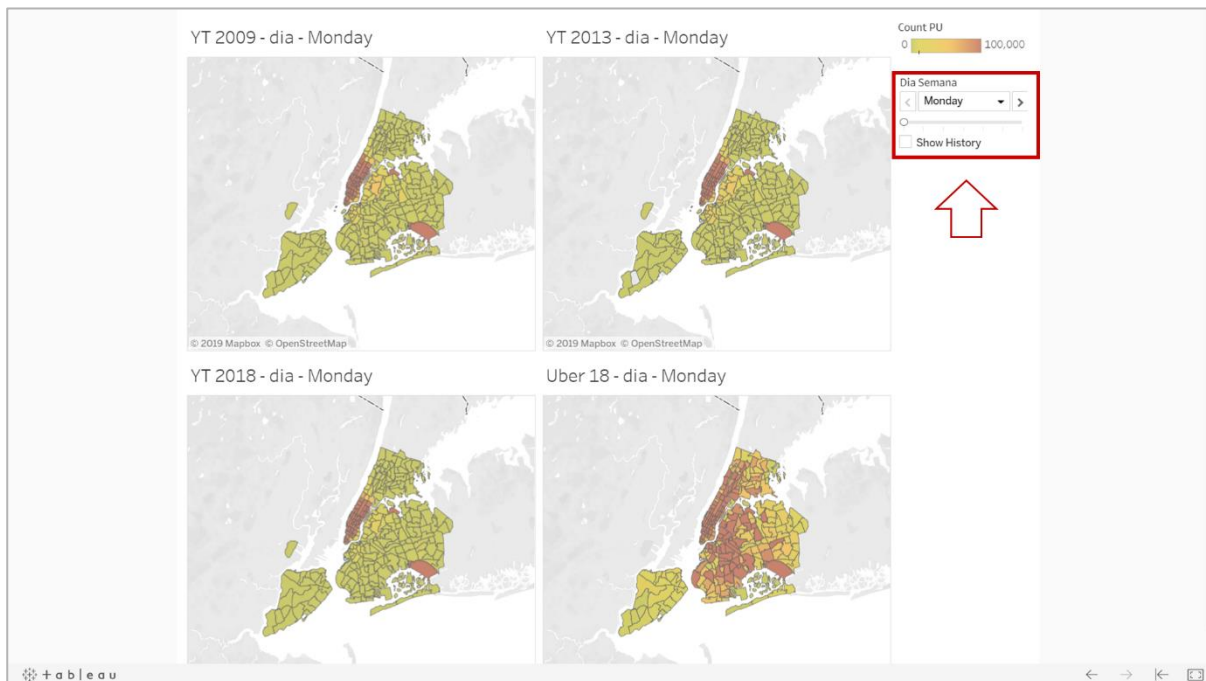


Figura 36: Indicação do botão utilizado para selecionar o dia da semana desejado na visualização dentro da página do Tableau Public com a análise do número de *pick-ups* de táxi por dia da semana.
Fonte: elaborada pelo autor.

Ao se observar as imagens, tanto para a variação entre os dias da semana quanto para a entre as horas de um dia, não há diferenças significativas entre as localizações preferidas para viagens de táxi ao longo do tempo, sejam elas de *pick-up* ou *drop-off*.

Adicionalmente, ressalta-se que, apesar dos números de táxi não terem apresentado grandes variações entre os anos analisados, eles são bastante diferentes dos de Uber. As corridas de táxi são quase todas concentradas dentro do *borough* de Manhattan, com outros lugares pontuais de concentração, como o aeroporto JFK no lado direito dos mapas, especialmente no caso do *pick-up* (o *drop-off* é ligeiramente mais espaçado, com corridas possivelmente saindo de Manhattan e indo para outros *boroughs*). Já as corridas de Uber têm uma capilaridade muito maior, atingindo zonas de táxi em outros *boroughs* e apresentam pontos de concentração relevantes em locais com baixa presença de táxi.

4.4.4. Localização de *pick-up* e *drop-off*

A localização de *pick-up* e *drop-off* das corridas podem ser visualizadas através dos diagramas de Sankey a seguir, feitos no Tableau. As figuras 37, 38 e 39 se referem às corridas de táxi em 2009, 2013 e 2018, respectivamente; e a figura 40 mostra as corridas de Uber em 2018. Nesses diagramas, a barra do lado esquerdo representa o percentual total de *pick-ups* de cada *borough* e a barra do lado direito os *drop-offs*. Os elos entre as barras são os fluxos que acontecem entre cada *borough*, de modo que os elos começam e possuem a cor dos *boroughs* de início das corridas e vão até o *borough* do final das corridas. A largura de cada elo é diretamente proporcional a quantidade de corridas entre as regiões.

Na figura 37, 93,5% dos *pick-ups* ocorreram em Manhattan e 4,6% em Queens. Os outros *boroughs* foram omitidos por terem números pouco expressivos. Os elos azuis são as corridas que partiram de Manhattan e a espessura dos mesmos indica a intensidade com a qual as corridas seguem o trajeto da região inicial para a final. A grande maioria das corridas iniciadas neste *borough* são direcionadas para ele mesmo, com uma parcela menor indo para Queens, outra para o Brooklyn, uma ainda menor para o Bronx e quantidades bem pouco expressivas para Staten Island e EWR. Também se observa que a maioria das corridas de Queens foi direcionada ou para Manhattan ou para a própria região. Além disso, a coluna dos *drop-offs* apresenta variação relevante em relação a dos *pick-ups*. Tanto Brooklyn quanto Queens ganharam um percentual maior de participação nas corridas, chegando a 4,2% e a 5,0%, respectivamente. Já Manhattan teve uma redução em seu número, passando de 93,5% para 90,1%.

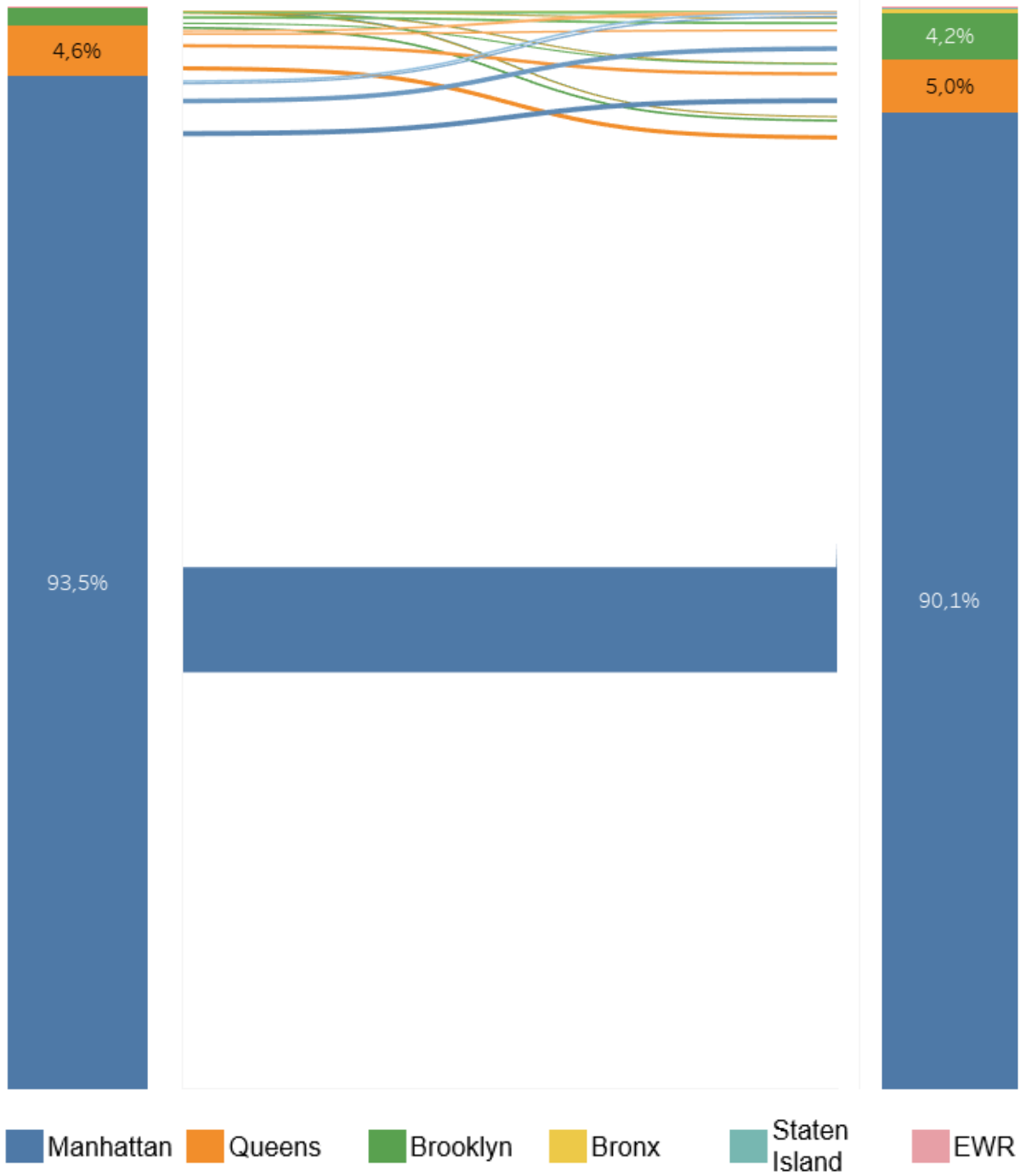


Figura 37: Diagrama de Sankey para as corridas de táxi em 2009, representando a origem e o destino das viagens entre os *boroughs* de Nova Iorque.
Fonte: elaborada pelo autor com dados da *TLC*.

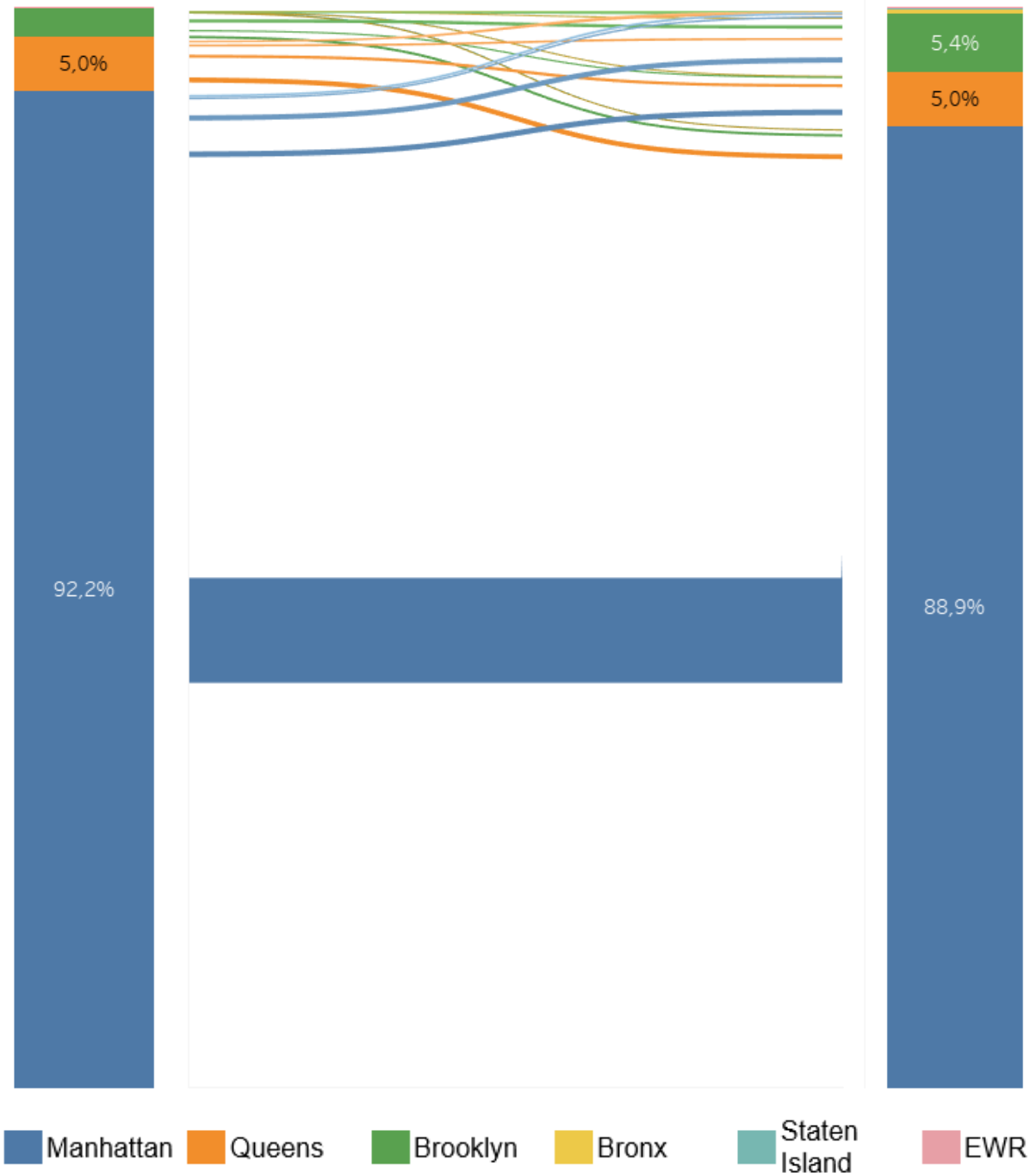


Figura 38: Diagrama de Sankey para as corridas de táxi em 2013, representando a origem e o destino das viagens entre os *boroughs* de Nova Iorque.
Fonte: elaborada pelo autor com dados da TLC.

Na figura 38, temos um cenário bastante parecido com o de 2009. A grande maioria das corridas ainda está concentrada em Manhattan, com um leve aumento do número de *pick-ups* e *drop-offs* nos outros *boroughs*. Destaca-se que houve uma variação no percentual de corridas iniciadas em Queens, mas o número de terminos ficou estável. Já o Brooklyn teve um aumento em ambas as dimensões.

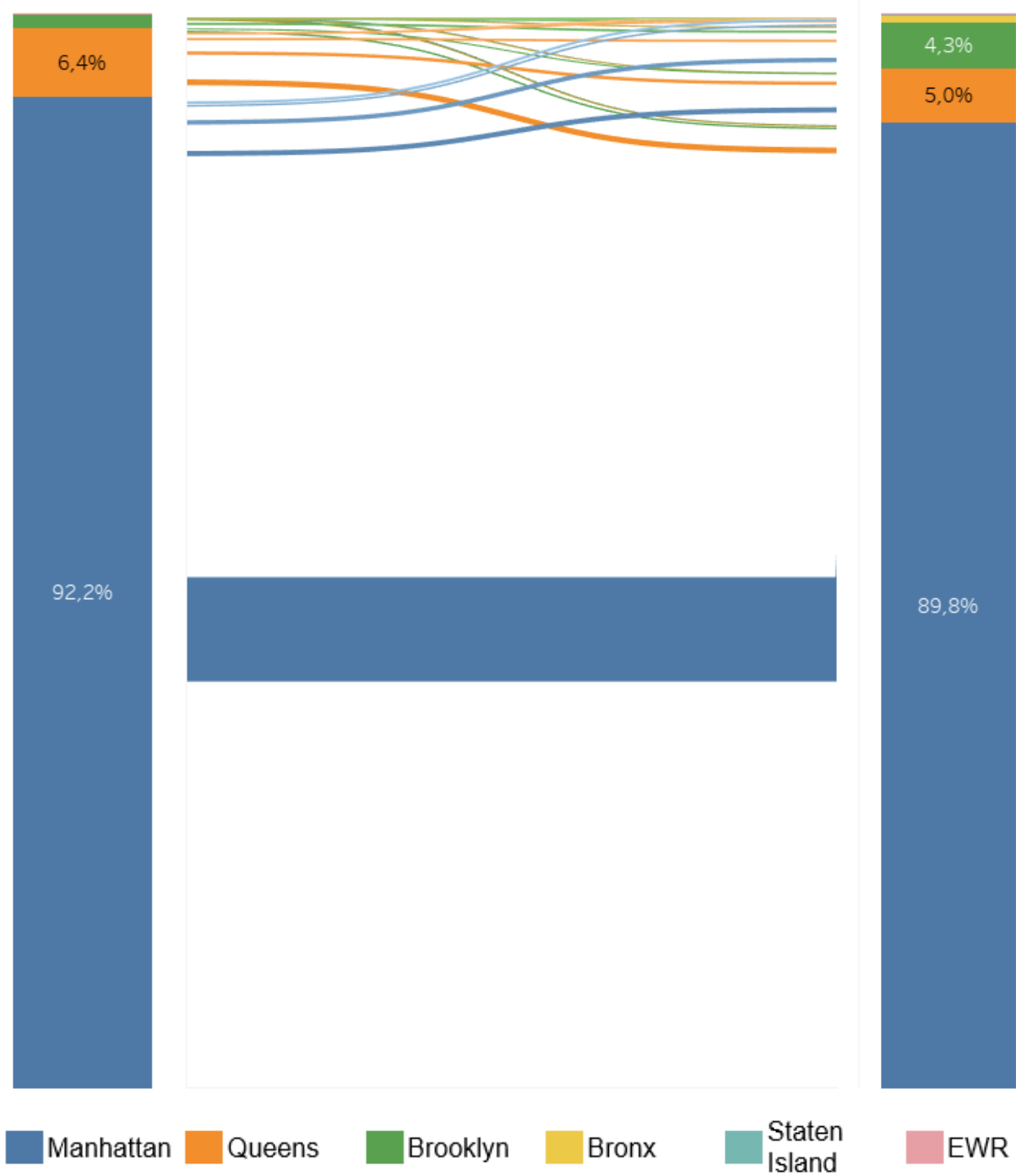


Figura 39: Diagrama de Sankey para as corridas de táxi em 2018, representando a origem e o destino das viagens entre os *boroughs* de Nova Iorque.
Fonte: elaborada pelo autor com dados da *TLC*.

Na figura 39, voltamos para um cenário mais parecido com o da figura 37, com Manhattan mais forte. Além disso, há uma queda expressiva de *drop-offs* em relação aos *pick-ups* em Queens, passando de 6,4% para 5%, e o Brooklyn figura novamente como mais relevante no destino do que no início das corridas.

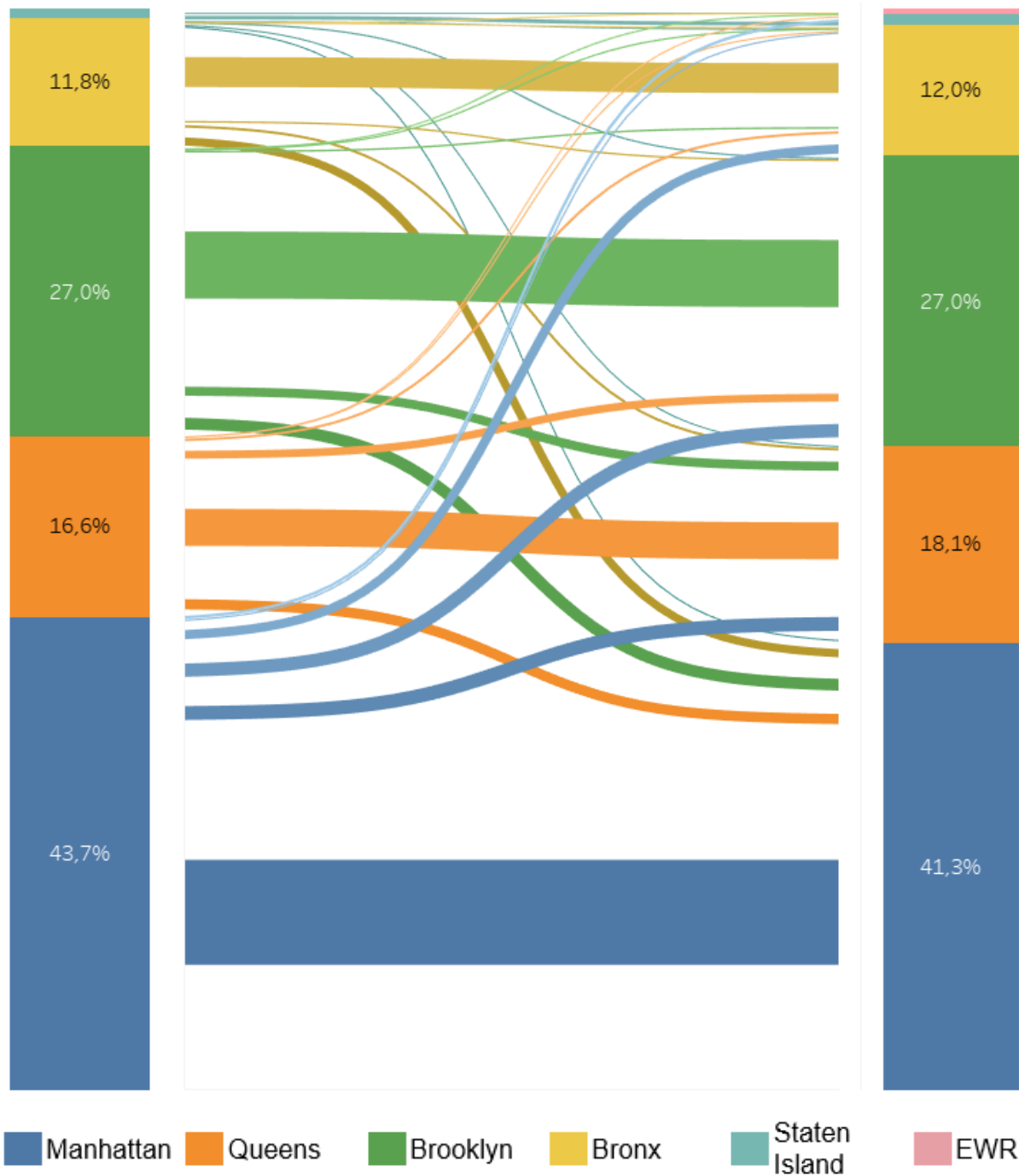


Figura 40: Diagrama de Sankey para as corridas de Uber em 2018, representando a origem e o destino das viagens entre os *boroughs* de Nova Iorque.
 Fonte: elaborada pelo autor com dados da TLC.

Na figura 40, pode-se observar que a distribuição de corridas da Uber é bastante diferente da vista nos diagramas dos táxis. Manhattan continua o bairro mais acessado, mas Queens, Brooklyn e Bronx também aparecem com bastante destaque, somando 55,4% dos *pick-ups* e 57,1% dos *drop-offs* no período, com cada uma possuindo mais de 10% das viagens em ambas as dimensões. Além disso, apesar dos elos entre os mesmos *boroughs* sempre serem os

mais largos, indicando forte presença de corridas nas mesmas regiões, os elos ligando localidades diferentes tem espessura maior do que os vistos nas corridas de táxi, indicando que os trajetos realizados pela Uber tendem a cruzar mais regiões da cidade. Dessa maneira, o diagrama indica que a empresa realiza corridas diferentes das que são realizadas pelo concorrente público, atendendo uma nova procura que era subatendida com a presença exclusiva dos táxis.

Ressalta-se que, apesar de pequenas diferenças, a distribuição de corridas entre os *boroughs* de Nova Iorque ficou praticamente constante para as viagens de táxi em 2009, 2013 e 2018, de modo a não se perceber nenhuma diferença significativa na distribuição destas corridas desde a entrada da concorrente no mercado.

4.4.5. Evolução das corridas em Manhattan

4.4.5.1. Evolução das corridas que começam e/ou terminam em Manhattan

As corridas que começam e/ou terminam no *borough* estão representadas na figura 41. As corridas desta categoria representam entre 94% e 95% de todas as corridas de táxi nos anos analisados. Entre 2009 e 2013, o mercado total ficou praticamente estável. Já para 2018, houve um crescimento de 34% no número de corridas. Apesar de ser um valor alto, ainda é tímido quando comparado com o crescimento de 110% visto na figura 17, para o mesmo período. Além disso, dado que as corridas desta classificação continuam bastante representativas do total das viagens de táxi, percebe-se que a entrada de novos competidores foi responsável pela queda no número de corridas nesta região, mas que não houve um impacto na distribuição geográfica das corridas realizadas, como já visto nas seções 4.4.3. e 4.4.4.

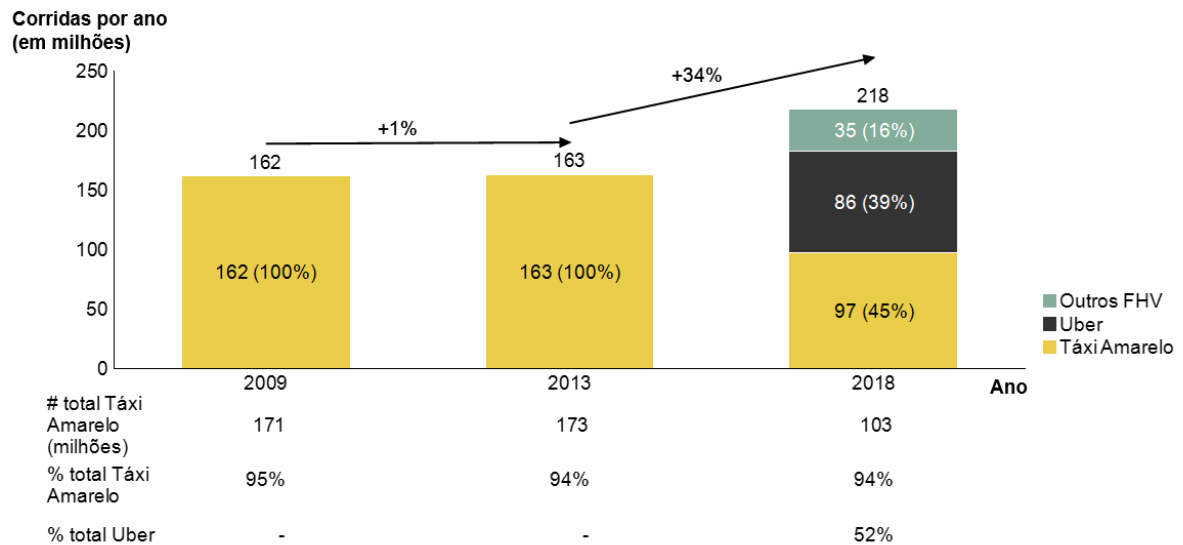


Figura 41: Evolução das corridas que começam e/ou terminam em Manhattan por tipo de transporte individual.

Fonte: elaborada pelo autor com dados da TLC.

4.4.5.1. Evolução das corridas que começam e terminam em Manhattan

Na figura 42, observa-se as corridas em que tanto o *pick-up* quanto o *drop-off* ocorreram dentro do *borough* de Manhattan. Nesse cenário, isola-se ainda mais a “fortaleza” do táxi, que é a região onde ocorre a maioria de suas viagens.

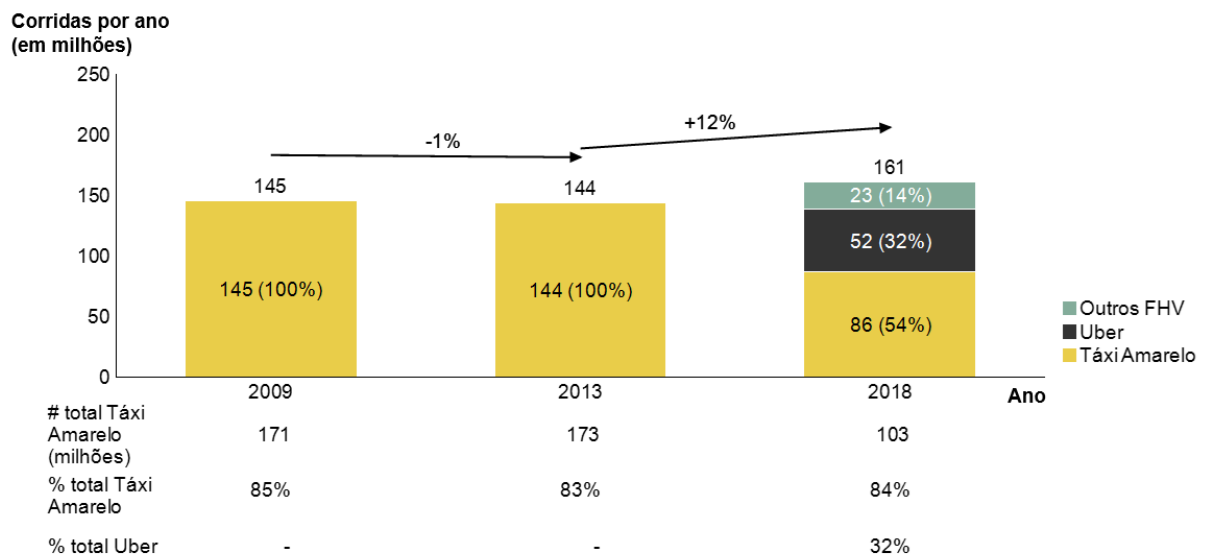


Figura 42: Evolução das corridas que começam e terminam em Manhattan por tipo de transporte individual.

Fonte: elaborada pelo autor com dados da TLC.

Nessa figura, o crescimento do número total de corridas é ainda menor quando comparado com a figura 17 (em comparação com a figura 41), com uma variação de apenas

12%, indicando que o mercado desse *borough* já é bastante saturado em termos de quantidade de corridas. Por conta disso, ao se observar que há pouco crescimento nas regiões nas quais a presença do táxi já era forte, pode-se afirmar que o aumento do número de corridas observado na figura 17 ocorre devido ao estímulo de demanda da oferta de *FHVs* em regiões nas quais a presença de táxi era modesta, e que esse aumento não ocorre de maneira linear em todas as regiões.

Ressalta-se que, mesmo com a diminuição do número de corridas de táxi de 2013 para 2018, o percentual de corridas que começam e terminam em Manhattan cresceu entre esses anos, passando de 83% para 85%, contra apenas 32% da Uber.

Dessa forma, entende-se que a Uber e outros meios de corrida *FHV* não apenas tem uma alta capacidade de geração de demanda em áreas nas quais os táxis atuam pouco, como também são uma concorrência forte para os táxis nas regiões em que eles já possuem presença forte.

5. CONCLUSÃO

A partir das análises apresentadas, conclui-se que a entrada da Uber no mercado de transporte individual da cidade de Nova Iorque não teve impacto expressivo nos padrões de deslocamento estudados para os táxis, sendo eles: (1) a distância e a duração das corridas; (2) a sazonalidade entre os dias da semana e entre os meses do ano; (3) as localizações de *pick-up* e *drop-off* das corridas, quando analisadas pelas variáveis horas do dia e por dias da semana; e (4) os fluxos de tramitação entre os *boroughs*.

Entretanto, apesar de não haver diferenças na forma de uso dos táxis, há duas situações em que a presença deste competidor privado impacta na mobilidade dos usuários deste transporte.

A primeira delas é o aumento da competitividade de regiões nas quais a presença de táxi já era forte. Na análise da evolução de corridas em Manhattan, mostra-se que houve um aumento tímido do número de corridas na região (apenas 12% para corridas que começam e terminam no *borough*), mas que a Uber foi capaz de absorver uma grande parte da demanda que já existia e já era atendida (32% do total de corridas, incluindo o aumento do mercado), concorrendo profundamente com os *Yellow Taxis*. Essa alteração, conforme Wallsten (2015), já mencionado na seção de justificativa (1.3.), traz benefícios por meio do aumento de qualidade do serviço prestado e está associado à redução nas reclamações dos consumidores às viagens de táxis, incluindo menor índice de defeitos em máquinas de cartão de crédito, menor incidência de ar condicionado e aquecimento quebrados, e redução de práticas inadequadas no processo de direção dos veículos.

A segunda mudança, observada nas localizações de *pick-up* e *drop-off* das corridas de Uber, é a geração de demanda por este tipo de transporte em áreas pouco servidas pelos táxis. Zonas localizadas fora de Manhattan, consistentemente ao longo do histórico estudado, foram subatendidas. Entretanto, a entrada dos *FHVs* no mercado permitiu que essas regiões tivessem acesso ao transporte privado individual, dobrando em 5 anos a quantidade total de corridas realizadas na cidade (figura 17). Essa modificação também é positiva pois facilita a locomoção dos novos usuários que antes não possuíam essa alternativa de transporte.

Além disso, pode-se perceber que há diferenças significativas entre os perfis das viagens realizadas por táxis e por veículos da Uber, que além de atingirem zonas de diferentes, tendem a ter tempos de duração maiores, provavelmente por transitar entre diferentes regiões. Estas conclusões estão em linha com o estudo de McKenzie e Baéz (2016), que diz que o táxi e a Uber possuem públicos distintos.

5.1. Trabalhos futuros

Para próximos estudos, recomenda-se entender com maior profundidade as razões pelas quais as corridas de táxi em Nova Iorque são concentradas em Manhattan e analisar o impacto da entrada de concorrentes privados no mercado de transporte público individual em cidades que podem ou não possuir a mesma concentração espacial de viagens.

Em especial, São Paulo é uma localidade interessante para ser estudada. A cidade possui similaridades com seu par estadunidense por se tratar de um grande centro urbano e também possuir uma grande concentração e centralização de suas atividades urbanas (Lencioni, 2008). Além disso, houve um aumento considerável no número de viagens de transporte público individual e privado no local, passando de 90,7 mil viagens diárias em 2007 para 468,4 mil em 2017, segundo a Pesquisa Origem e Destino 2017 (Companhia do Metropolitano de São Paulo – Metrô, 2017). O estudo também ressalta que a modalidade de táxi convencional, sem o uso de aplicativos, foi responsável por 98,4 mil corridas diárias na data mais recente da pesquisa, representando um crescimento de 8,4% em relação ao primeiro ano – diferentemente do caso de Nova Iorque, em que há queda na quantidade de viagens realizadas por *Yellow Cabs*. Entretanto, a cidade não possui informações sobre estes meios de transporte em suas políticas de dados abertos, o que dificultaria a realização do trabalho.

Outra sugestão é explorar como eram realizados os deslocamentos de usuários que hoje são atendidos pela Uber e que antes não realizavam viagens de transporte público individual, de modo a entender quais meios de transporte que foram substituídos por esta alternativa de locomoção. Dessa maneira, poderia compreender-se melhor o impacto do surgimento dos concorrentes privados na mobilidade e no trânsito de uma cidade, verificando se eles atuam de maneira complementar ou concorrente ao transporte público coletivo.

REFERÊNCIAS BIBLIOGRÁFICAS

- ABEPRO. **Áreas da Engenharia de Produção**. Disponível em: < <http://portal.abepro.org.br/a-profissao/>>. Acesso em 31 de outubro de 2019.
- ALTERYX. **Alteryx Designer 11.0**. Version 2018.4.5.55178. Estados Unidos. Alteryx, Inc., 2018. Disponível em: <<https://www.alteryx.com/resources/product-support/download-alteryx>>. Acesso em 20 de março de 2019.
- ANGRIST, J.; CALDWELL, S.; HALL, J. **Uber vs. Taxi: A Driver's Eye View**. NBER Working Paper #23,891, 2017.
- BONOMO, R. **Metodologia da Pesquisa Científica**. IFSC-SBI. 2007.
- CALABRESE, F.; DIAO, M.; LORENZO, G.D.; FERREIRA, J.; RATTI, C. **Understanding individual mobility patterns from urban sensing data: a mobile phone trace example**. Transp. Res. Part C, 26 (January), p. 301-313, 2013.
- Cannon, S.; Summers, L.H. **How Uber and the sharing economy can win over regulators**. Harvard Business Review, 2014.
- CHU, D.; SHEETS, D.A.; ZHAO, Y.; WU, Y.; YANG, J.; ZHENG, M.; CHEN, G. **Visualizing hidden themes of taxi movement with semantic transformation**. Visualization Symposium (PacificVis), 2014 IEEE Pacific, IEEE, p. 137-144, 2014.
- COMPANHIA DO METROPOLITANO DE SÃO PAULO – METRÔ. **Pesquisa Origem e Destino**. 2017. Disponível em: < <http://www.metro.sp.gov.br/pesquisa-od/>>. Acesso em 07 de novembro de 2019.
- CRAMER, J.; KRUEGER, A.B.; **Disruptive change in the taxi business: the case of Uber**. Am. Econ. Rev., 106 (5), pp. 177-182, 2016.
- DEMPSEY, P. **Taxi Industry Regulation, Deregulation & Reregulation: The Paradox of Market Failure**, 24 TRANSP. L. J. vol. 24:73 p. 73-120, 1996.
- DERI, J. A.; MOURA, J. M. F. **Taxis in New York City: A network perspective**. Proc. 49th Asilomar Conf. Signals, Syst., Comput., p. 1829–1833, 2015.
- DORAISWAMY, H.; ZACHARATOU, E. T.; MIRANDA, F.; LAGE, M.; AILAMAKI, A.; SILVA, C. T.; FREIRE, J. **Interactive visual exploration of spatio-temporal urban data sets using urbane**. SIGMOD, 2018.
- EREVELLES, S.; FUKAWA, N.; SWAYNE, L. **Big data consumer analytics and the transformation of marketing**. Journal of Business Research, 69, p. 897-904, 2016.
- FERREIRA, N.; POCO, J.; VO, H.T.; FREIRE, J.; SILVA, C.T. **Visual exploration of big spatio-temporal urban data: a study of New York City taxi trips**,

- GALLICK, E.; SISK, D. **Reconsideration of Taxi Regulation**, Journal of Law, Economics & Organization, vol. 3, no. 1, p. 117-128, 1978.
- GANDOMI, A.; HAIDER, M. **Beyond the hype: Big data concepts, methods, and analytics**. International Journal of Information Management, p. 137-144, 2015.
- GONZÁLEZ, M.C.; HIDALGO, C.A.; BARABÁSI, A.-L. **Understanding individual human mobility patterns** Nature, 453, p. 779-782, 2008.
- GUNASEKARAN, A.; PAPADOPOULOS, T.; DUBEY, R.; WAMBA, S.F.; CHILDE, S.J.; HAZEN, S. B. **Big data and predictive analytics for supply chain and organizational performance**. Journal of Business Research, 70, p. 308-317, 2017.
- HERRING, R.; HOFLEITNER, A.; ABBEEL, P.; Bayen, A. **Estimating arterial traffic conditions using sparse probe data**. Paper presented at the Intelligent Transportation Systems (ITSC), 2010 13th International IEEE Conference on., 2010.
- HOQUE, M. A.; HONG, X.; DIXON, B. **Analysis of mobility patterns for urban taxi cabs**. Paper presented at the Computing, Networking and Communications (ICNC), 2012 International Conference on., 2012.
- JIANG, S.; FERREIRA JR., J; GONZÁLEZ, M.C. **Activity-based human mobility patterns inferred from mobile phone data: a case study of Singapore**. In: UrbComp'15, Sydney, Australia, 2015.
- KANG, C.; MA, X.; TONG, D.; LIU, Y.; **Intra-urban human mobility patterns: An urban morphology perspective**. Physica A, 391 (4), p. 1702-1717, 2011.
- KHOURY, M.J.; IOANNIDIS, J.P.A. **Big data meets public health**. Science, 346 p. 1054-1055, 2014.
- LANEY, D. **3-D data management: Controlling data volume, velocity and variety**.
- LEE, J. **Analysis on the waiting time of empty taxis for the taxi telematics system**. ICCIT, 2008
- LENCIONI, S. **Concentração e centralização das atividades urbanas: uma perspectiva multiescalar. Reflexões a partir do caso de São Paulo**. Rev. Geogr. Norte Gd. n. 39, p. 7-20, 2008.
- LI, J.; TAO, F.; CHENG, Y.; ZHAO, L.; **Big Data in product lifecycle management**. The International Journal of Advanced Manufacturing Technology. 2015.
- LIU, J.; LI, J.; LI, W.; WU, J. **Rethinking big data: A review on the data quality and usage issues**, ISPRS J. Photogram. Remote Sens., vol. 115, p. 134–142, 2016.
- LIU, X.; GONG, L.; GONG, Y.; LIU, Y. **Revealing travel patterns and city structure with taxi trip data** J. Transp. Geogr., 43, p. 78-90, 2015.

- LIU, Y.; KANG, C.; GAO, S.; XIAO, Y.; TIAN, Y. **Understanding intra-urban trip patterns from taxi trajectory data**. J. Geogr. Syst., 14 (4), p. 463-483, 2012.
- LV, Y.; DUAN, Y.; KANG, W.; LI, Z.; WANG, F.-Y. **Traffic flow prediction with big data: A deep learning approach**. Intelligent Transportation Systems, IEEE Transactions on, vol. PP, no. 99, p. 1–9, 2014.
- MCAFEE, A., & BRYNJOLFSSON, E. **Big Data: The Management Revolution**. Harvard Business Review, 2012.
- MCKENZIE, G.; BAÉZ, C. **Uber vs. taxis: Event detection and differentiation in New York City**. International Conference on GIScience Short Paper Proceedings; UC Berkeley: Berkeley, CA, USA, 2016.
- MEKKO GRAPHICS. **Mekko Graphics**. Version 9.7.2686. Estados Unidos. Knowledge Management Associates, LLC, 2018. Disponível em: <<https://www.mekkographics.com/resources/get-started/download/>>. Acesso em 19 de fevereiro de 2018.
- MICROSOFT. **Excel 2016 MSO 64-bit**. Version 16.0.4849.1000. Estados Unidos. Microsoft Corporation, 2016a. Disponível em: <<https://products.office.com/en/excel>>. Acesso em 19 de fevereiro de 2018.
- MICROSOFT. **PowerPoint 2016 MSO 64-bit**. Version 16.0.4849.1000. Estados Unidos. Microsoft Corporation, 2016b. Disponível em: <<https://products.office.com/en/powerpoint>>. Acesso em 19 de fevereiro de 2018.
- MIGUEL, P.; FLEURY, A.; MELLO, C.; NAKANO, D.; LIMA, E.; TURRITONI, J.; HO, L.; MORABITO, R.; MARTINS, R.; SOUSA, R.; COSTA, S.; PUREZA, V. **Metodologia de pesquisa em engenharia de produção e gestão de operações**. 2ª ed. Rio de Janeiro: Elsevier. 2012.
- MURDOCH, T.B.; DETSKY, A.S. **The inevitable application of big data to health care**. J. Am. Med. Assoc., 309 (13), p. 1351-1352, 2013.
- NEW YORK TIMES. **Uber I.P.O. Values Ride-Hailing Giant at \$82.4 Billion**. 9 mai. 2019. Disponível em: <<https://www.nytimes.com/2019/05/09/technology/uber-ipo-stock-price.html>>. Acesso em 10 de setembro de 2019.
- NEW YORK TIMES. **Where Yellow Cabs Didn't Go, Green Cabs Were Supposed to Thrive. Then Came Uber**. 3 set. 2018. Disponível em: <<https://www.nytimes.com/2018/09/03/nyregion/green-cabs-yellow-uber.html>>. Acesso em 10 de setembro de 2019.
- NIE, Y.M. **How can the taxi industry survive the tide of ridesourcing? Evidence from Shenzhen, China**. Transp. Res. Part C: Emerg. Technol., 79, pp. 242-256, 2017.

OKAMOTO, K. **What is being done with open government data? An exploratory analysis of public uses of New York City open data.** Webology 13, n. 1: 1-12, 2016.

PENG, C.; JIN, X.; WONG, K.-C.; SHI, M.; LI, P. **Collective human mobility pattern from taxi trips in urban area.** PLoS One, 7, p. e34487, 2012.

POULSEN, L.; DEKKERS, D.; WAGENNAR, N; SNIJDERS, W.; LEWINSKY, B.; MUKKAMALA, R.; VATRAPU, R. **Green cabs vs. Uber in New York City.** Proceedings of the 2016 IEEE International Congress on Big Data, p. 222-229, 2016.

NEW YORK CITY DEPARTMENT OF CITY PLANNING. **Population - Current and Projected Populations.** 1 jul. 2018. Disponível em: <<https://www1.nyc.gov/site/planning/planning-level/nyc-population/current-future-populations.page>>. Acesso em 14 de outubro de 2019.

SONG, C.; QU, Z.; BLUMM, N.; BARABÁSI, A.-L. **Limits of predictability in human mobility.** Science, 327, p. 1018-1021, 2010.

TABLEAU SOFTWARE. **Tableau Desktop Personal Edition 2019.2 64-bit.** Version 20192.19.0515.1630. Estados Unidos. Tableau Software, Inc., 2019. Disponível em: <<https://www.tableau.com/products/desktop>>. Acesso em 20 de junho de 2019.

TLC. **TLC Trip Records User Guide.** 18 out. 2018. Disponível em: <https://www1.nyc.gov/assets/tlc/downloads/pdf/trip_record_user_guide.pdf>. Acesso em: 18 de maio de 2019.

UBER. **Prospectus filed pursuant to Rule 424(b)(4).** 9 mai. 2019. Disponível em: <<https://d18rn0p25nwr6d.cloudfront.net/CIK-0001543151/c4e8579e-f8c4-469c-b86f-20245606c8a3.pdf>>. Acesso em 10 de setembro de 2019.

UBER. **Uber NYC has launched.** 4 mai. 2011. Disponível em: <<https://www.uber.com/blog/new-york-city/uber-nyc-launches-service/>>. Acesso em 19 de maio de 2019

VELOSO, M.; PHITHAKKITNUKON, S.; BENTO, C.; FONSECA, N.; OLIVIER, P. **Exploratory study of urban flow using taxi traces.** Paper presented at the First Workshop on Pervasive Urban Applications (PURBA) in conjunction with Pervasive Computing, San Francisco, California, USA, 2011.

WALLSTEN, S. **The Competitive Effects of the Sharing Economy: How is Uber Changing Taxis?** New York: Technology Policy Institute, 2015.

WANG, G.; GUNASEKARAN, A.; NGAI E.W.; PAPADOPOULOS, T. **Big data analytics in logistics and supply chain management: certain investigations for research and applications.** International Journal of Production Economics, 176 p. 98-110, 2016.

ZHAO, K.; MUSOLESI, M.; HUI, P.; RAO, W.; TARKOMA, S. **Explaining the power-law distribution of human mobility through transportation modality decomposition.** Nature Scientific Reports, vol. 5, no. 9136, 2015.

ZHENG, Y.; LUI, F.; HSIEH, H.-P. **U-Air: when urban air quality inference meets big data.** Proceedings of the 19th ACM SIGKDD International Conference on Knowledge Discovery and Data Mining, ACM, Chicago, IL, USA, p. 1436-1444, 2013.

ZHENG, X.H.; CHEN, W.; WANG,P.; SHEN, D.Y.; HEN, S.H.; WANG, X.; ZHANG, Q.P.; YANG, L.Q. **Big data for social transportation.** IEEE Trans. Intell. Transp., 17, p. 620-630, 2016.