How Does Viral Infection Affect Taxi Service Reliance?

MAST30034 Assignment 1

Xavier Travers
Student ID: 1178369
TODO: Github Repository

August 11, 2022

1 Introduction

Viral infection has been on everyone's mind in the past few years due to the COVID-19 pandemic. With lockdowns and fears of infection, it is a natural assumption that many people-facing industries such as ride-hailing have suffered in demand. To what extent is such an assumption true? This report aims to contribute to a body of works attempting to quantify the effect that larger-scale viral diseases can have on different industries. Specifically investigated will be the effects that influenza/COVID-19 case rates have on taxi trip rates and passenger counts.

This report will detail the statistical analysis involved in revealing the extent to which changes in case rates of a viral infection may affect taxi trip rates and passenger counts.

1.1 Datasets

- The New York City Taxi and Limousine Commission (TLC) provides a dataset of taxi service trips which captures information such as type of taxi, travel distance, general pickup/dropoff locations and driver-input passenger counts [1]. Throughout this report, this is referred to as the TLC dataset. In this report, the focus is placed on the dataset concerning New York's Yellow hail taxis. High frequency hired vehicle service (e.g. Uber) data is not included in analysis, since the dataset is too recent for the purposes of this report (starting only in early 2019). Each entry in this dataset contains several values, but this report only focuses on trip distance, passenger count, date of pickup, and the pickup/dropoff locations.
- Influenza case rates are recorded on a weekly basis by the New York Department of Health [2]. Case rates in this dataset are dated based on Morbidity and Mortality Weekly Report (MMWR) weeks, which are generated using rules defined by the CDC [3]. Each entry in this dataset contains an MMWR week, county (within the state of New York), type of Influenza (A, B or unspecified), and case count. The dataset provides coverage over the whole selected timeline of data.
- COVID-19 case rates have been recorded daily by the New York Department of Health and Mental Hygiene [4]. This dataset begins on the last day of february, when the first official cases of COVID-19 were recorded in New York City. Each entry in this data set contains a date and several of the daily COVID-19 rates by borough (e.g. count of hospitalizations on the day in the Bronx). Of specific interest is the daily case count per borough.

1.2 Timeline

The overall timeline considered for the data spans from the start of 2018 to the end of 2021. This can be split into two shorter timelines of equal length: before COVID-19, and after COVID-19. Data from 2022 is not included, since the COVID-19 dataset is automated and has an increased likelihood of containing false information with newer data. Data from before 2018 is not included for the sake of symmetry and to reduce code runtime.

2 Method

2.1 Preprocessing

2.1.1 Cleaning

TLC Dataset

1. test

2.1.2 Necessary Extracted Features

The flu dataset contains detail only on a weekly basis, while the other datasets used contain daily data. This means that the data needs to be aggregated by MMWR week value, which allows the construction of linear models on weekly data. While the granularity of the data suffers, the large timeline ensures that at least 52 points of data are available for analysis per time period. Also necessary for connecting viral infection case rate datasets to the taxi datasets are the extraction of borough.

2.1.3 Aggregation

All data is aggregated by MMWR week. The datasets are then all joined on the MMWR week, and corresponding boroughs. Importantly, for each MMWR week, two sets of case rates are recorded: those corresponding to the pickup borough, and those corresponding to the dropoff borough. For each MMWR week aggregate, total case rates are recorded, as well as total average trip distances and average passenger counts.

2.2 Analysis and Modelling

2.3 Geospatial Visualisation

3 Recommendations

4 Conclusions

References

[1] New York City Taxi and Limousine Commission. *TLC Trip Record Data*. https://www1.nyc.gov/site/tlc/about/tlc-trip-record-data.page. Accessed: 2022-08-06.

[2] New York State Department of Health. *Influenza Laboratory-Confirmed Cases By County: Beginning 2009-10 Season*. https://health.data.ny.gov/Health/Influenza-Laboratory-Confirmed-Cases-By-County-Beg/jr8b-6gh6. Accessed: 2022-08-09.

[3] CDC. *MMWR Weeks*. https://ndc.services.cdc.gov/wp-content/uploads/MMWR_Week_overview.pdf. Accessed: 2022-08-09.

[4] Department of Health and Mental Hygiene (DOHMH). *COVID-19 Daily Counts of Cases, Hospitalizations, and Deaths*. https://data.cityofnewyork.us/Health/COVID-19-Daily-Counts-of-Cases-Hospitalizations-an/rc75-m7u3. Accessed: 2022-08-09.