**Vik Shah, David Cheung, Qasse Sakariya**

Project thesis

Graphical user interface

Description automatically generated with medium confidence

Table of Contents

[1 Project Module Vik Shah 3](#_Toc99193086)

[1.1 Introduction 3](#_Toc99193087)

[1.2 Weather & Uber-Pickups Data 3](#_Toc99193088)

[1.2.1 Month 3](#_Toc99193089)

[1.2.2 Day of the Week 4](#_Toc99193090)

[1.2.3 Weather Conditions 4](#_Toc99193091)

[1.3 Google Maps API Call 4](#_Toc99193092)

[1.3.1 Weather Conditions 4](#_Toc99193093)

[1.3.2 Uber Pickup Locations 5](#_Toc99193094)

[1.4 Conclusion 5](#_Toc99193095)

# Project Module Vik Shah

## Introduction

The analysis completed here is to identify external influences on uber pickups and advise uber based on the data how to optimize and enhance their revenue.

A few key hypotheses to analysis include:

1. Does the date in a month influence frequency of uber pickups e.g., school, or public holidays?
   1. **N0: Dates in a month decreases the frequency on uber pickups**
   2. N1: Dates in a month increases or has no effect on the frequency on uber pickups
2. Day of the week influence frequency of uber pickups e.g., weekends
   1. **N0: Saturdays increases the frequency on uber pickups**
   2. N1: Saturdays decreases or has no effect on the frequency on uber pickups
3. Weather conditions
   1. N0: Higher Temperature increases the frequency on uber pickups
   2. **N1: Higher Temperature decreases or has no effect on the frequency on uber pickups**
4. Location of pickups
   1. N0: Higher frequency of uber pickups occur with 5km of an airport
   2. **N1: Higher frequency of uber pickups occur not within 5km of an airport**

## Weather & Uber-Pickups Data

Approach to project assignment:

1. Obtain uber-pickups and weather data for a particular month and same city

* Selected July 2014 – ease of open-source data
* Select New York City – highly populated city for opportunity of big data

1. Merge the data sets and clean data to test the hypotheses

* Dropping null or NaN values to ensure a clean data set

1. Complete statistical analysis e.g. plotting graphs or calculating correlation coefficients

* Using the min/ mean/ describe functions to identify statistical significance

### Month

* Graphing the number of pickups per day over the course of the month
* Observing the impact of summer holidays, significant surge
* Observing the impact of fourth of July Public holiday, lowest pickups in the month

### Day of the Week

* Graphing the average number of uber pickups of the weekdays
* Given the data shows 5 Tue/Wed/Thu & 4 Fri/Sat/Sun/Mon in the month, the data was normalised to have the graph show the average pickups per weekday
* Observing that Saturdays have the highest pickups and Sundays the lowest
* Further investigation to be conducted as to why. Possible hypotheses include:
* Saturday Nights more people consume alcohol hence would require a ride home
* Sundays more people don’t go out, instead recover/ prepare for the work week ahead
* Tuesdays/ Wednesdays are the second highest, and often are cheaper to travel by flight
* If this last theory is possible, we would expect to see lots of pickups around the airports
* Let's see if the heatmap or map marker supports this hypothesis

### Weather Conditions

* Another influence on uber pickups may be the weather
* First the data needed to be cleaned i.e., using the first 13days or the holiday period of the month
* Plotting the Temperature vs. Number of Pickups as a scatter plot, allows for a correlation calculation to statistically prove of disprove that the hotter the temperature the more uber pickups occur
* The resulting infographic shows a correlation coefficient or r value = 0.394
* Suggesting this correlation is not statistically significant to state there is a relationship for this month between number of pickups and temperature

## Google Maps API Call

### Weather Conditions

* A similar weather assessment may be conducted on precipitation
* However, upon completing a google maps api call and plotting a heatmap for pickup locations vs. weighting of rain.
* The heatmap shows Manhattan New York as unaffected by rain during this particular month
* Hence not likely to have a strong correlation coefficient or statistically significant result between the relationship of Precipitation vs. Number of Pickups

### Uber Pickup Locations

* The final analysis is what to advise Uber, regarding where drivers should be positioned.
* Merge the data sets to identify epic centre of uber pickups per day
* Plotting the most frequent pickup locations on google maps. Key observations include:
* The epic centre of uber pickups for most days in this month are in the lower east side of Manhattan
* It is identified that there are no airports in that area, however there is a hospital which suggests or supports the hypothesis that taxi/ ubers are quicker than ambulances in NYC
* Additionally, nearby Hotels/ Parking Lots & Shopping Malls have been identified for uber drivers to wait in the most frequent pickup location spots around the city and plotted on a google maps figure with marker stating these locations.

## Conclusion

In conclusion the following observations were made, and hypothesis proven/ disproven for this month and year in NYC, rendering suggestions and advise to be shared with uber to increase revenue.

* Dates of the month:
* Summer Holiday do influence (increase) frequency on uber pickups
* Public Holidays (4th July) do influence (decrease) frequency on uber pickups
* Days of the week:
  + Saturdays do influence (increase) the frequency on uber pickups
  + Sundays do influence (decrease) the frequency on uber pickups
* Weather conditions:
  + Temperature doesn’t influence (no significant effect) on the frequency on uber pickups
  + Precipitation doesn’t influence (no significant effect) on the frequency on uber pickups
* Location of pickups
  + Airport location doesn’t influence (no significant effect) on the frequency on uber pickups
  + Hospital location suggests an influence (increase) on the frequency on uber pickups
  + Lower East Side Manhattan suggests an influence (increase) on the frequency on uber pickups locations