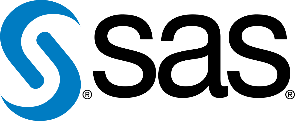
**

***PROJECT 2***

***Analysis on UBER Rides***

*Vaishnavi Ayyadurai*

*WB9452*

**INTRODUCTION**

This project is about extracting insights from UBER pickups/trips in New York City from Jan 2015 to June 2015. Also, other For-Hire Vehicles (FHV)trip data is being extracted, manipulated and compared to analyze the performance of UBER.

Raw data for this project is available in Kaggle:

<https://www.kaggle.com/fivethirtyeight/uber-pickups-in-new-york-city>

**PURPOSE**

The most congested parts of Manhattan are being overwhelmed by the addition of more than 2,000 new for-hire vehicles each month. They are facing the addition of over 25,000 cars to their streets over the next year -- the rough equivalent of two times the total number of yellow taxis in all of New York City.

The mayor is trying to “cap” Uber, Lyft, cap every service. What that means is it's a cap on jobs. It's killing over 10,000 jobs. It means poor service in underserved areas like the Bronx, like Queens.

So they are going to fight this with everything (facts-data) they have to substantiate the importance and usefulness of UBER trips for passengers, Number of trips per day/per month been availed by people at Bronx and different Base offices. And also how do they perform compared to other FHV vehicles.

**PROPOSAL**

We are going to compare the Total number of trips made by Other FHV vehicles and UBER for each Base\_numbers, for every month, for every location, number of vehicles used to compare the performance of UBER with other transportation companies. These summary data arrived will be used for high level business decision making, in order to improvise the business strategies on areas of weakness compared to other FHV’s.

This will also help them to argue with the Mayor with the relevant data so as to uncap UBER services compared to others.

We have used *PROC SQL, MACROS, PROC GPLOT* for our analysis.

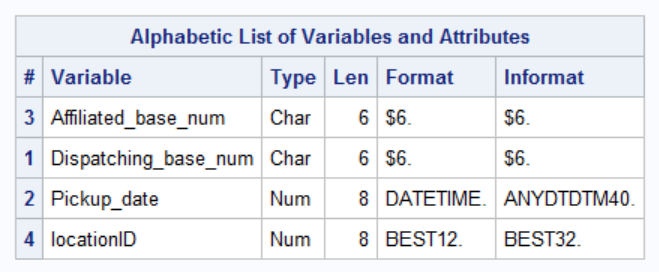
**Source data:**

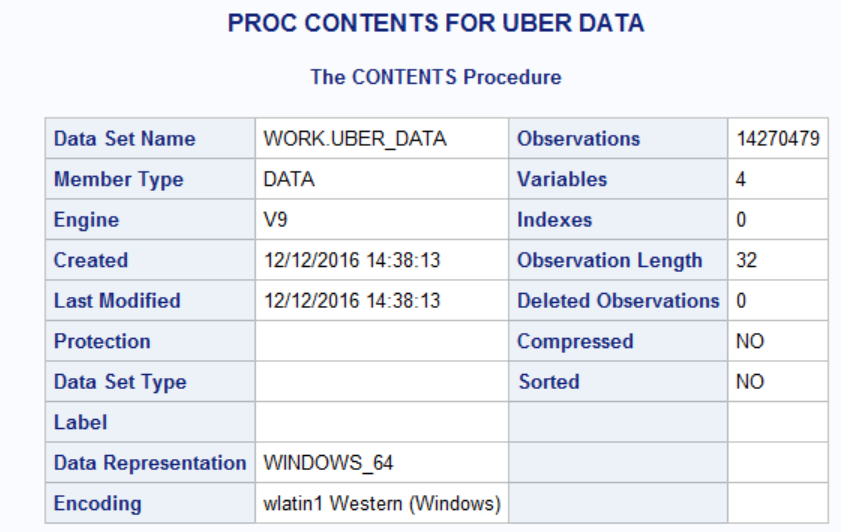
<https://www.kaggle.com/fivethirtyeight/uber-pickups-in-new-york-city/downloads/uber-pickups-in-new-york-city.zip>

**DATA:**

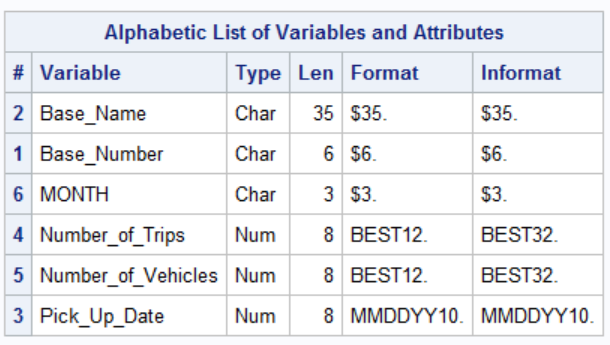
There are two raw data sets available:

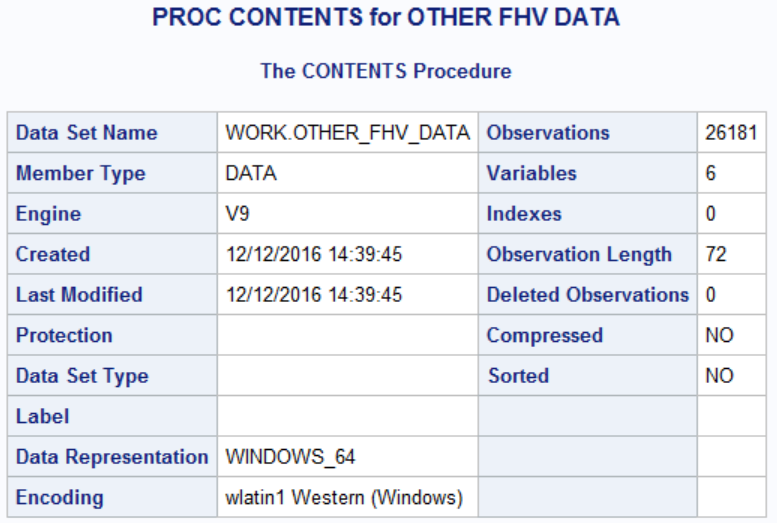
1. **UBER DATA:**





2.**OTHER FHV DATA:**





**PROCEDURE**

1. Firstly, missing values are removed in both the data sets by using below code:

/\*REMOVING MISSING VALUES\*/

**data** OTHER\_FHV\_DATA;

set OTHER\_FHV\_DATA;

if cmiss(of \_all\_) then delete;

**run**;

**data** UBER\_DATA;

set UBER\_DATA;

if cmiss(of \_all\_) then delete;

**run**;

1. Number of vehicles and trips for each base name for other FHV vehicles is as shown below

**/\*USING MACROS\*/**

%LET TRIPS=NUMBER\_OF\_TRIPS;

%LET VEHICLES=NUMBER\_OF\_VEHICLES;

TITLE "NO OF VEHICLES & TRIPS FOR EACH BASE NAME in other FHV ";

**PROC** **SQL**;

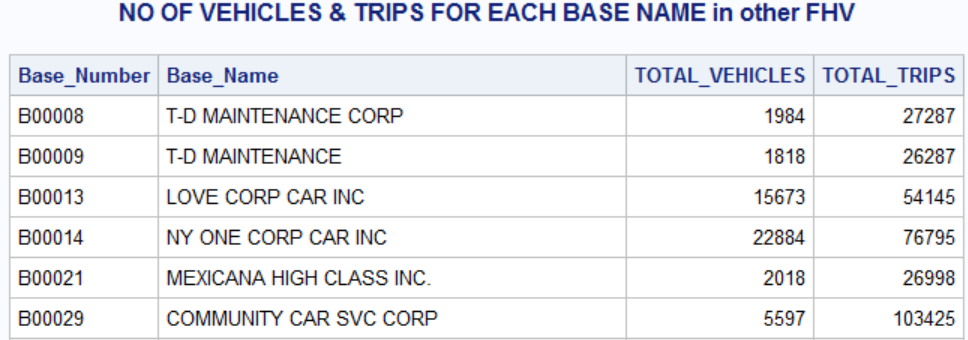
SELECT BASE\_NUMBER,BASE\_NAME, SUM(&VEHICLES)AS TOTAL\_VEHICLES ,SUM(&TRIPS)AS TOTAL\_TRIPS

from OTHER\_FHV\_DATA

group by **1**,**2**

order by **1**,**2**;

**QUIT**;

(Sample data only):

1. Number of Trips for every Month-Other FHV Vehicles:

TITLE "NO OF TRIPS FOR EVERY MONTH-OTHER-FHV";

**PROC** **SQL**;

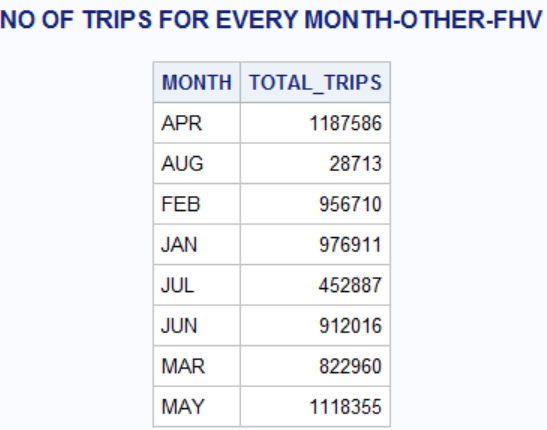
SELECT MONTH,SUM(&TRIPS)AS TOTAL\_TRIPS

FROM OTHER\_FHV\_DATA

GROUP BY **1**

ORDER BY **1**;

**QUIT**;



**Graphical Representation:**

TITLE "NO OF TRIPS BY MONTH";

pattern1 value=L1;

**proc** **gchart** data=OTHER\_FHV\_DATA;

vbar MONTH / sumvar=&TRIPS

type=SUM;

**run**;

**quit**;



1. **Number of trips for every base for every month:**

TITLE "NO OF TRIPS FOR EVERY MONTH FOR EVRY BASE\_NUMBERS -OTHER-FHV";

**PROC** **SQL**;

SELECT BASE\_NUMBER,MONTH,SUM(NUMBER\_OF\_TRIPS)AS TOTAL\_TRIPS

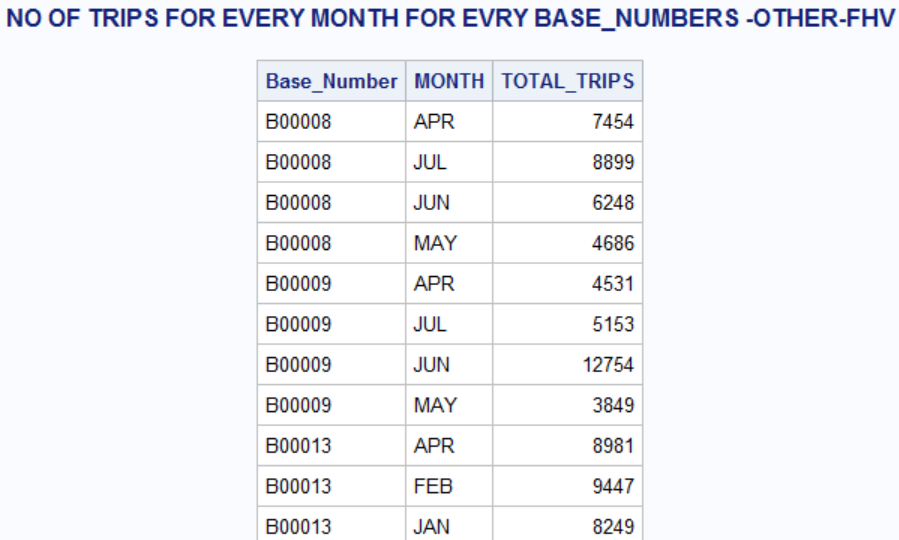
FROM OTHER\_FHV\_DATA

GROUP BY **1**,**2**

ORDER BY **1**,**2**;

**QUIT**;

**Sample data:**



**UBER DATA ANALYSIS:**

1. **Total Trips for every UBER Base:**

%LET BASE\_NO=Dispatching\_base\_num;

TITLE "NO OF TRIPS FOR EACH UBER BASE ";

**PROC** **SQL**;

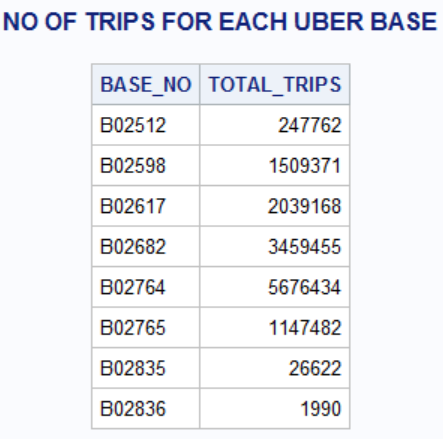
SELECT &BASE\_NO AS BASE\_NO,COUNT(&BASE\_NO)AS TOTAL\_TRIPS

from UBER\_DATA

group by **1**

order by **1**;

**QUIT**;

****

1. **Total Trips for Every Month:**

/\*Extracting month from the pipckup date\*/

**data** uber\_data;

set uber\_data;

Month=SUBSTR(put(Pickup\_date,DATETIME16.),**3**,**3**);

**run**;

TITLE"NUMBER OF TRIPS FOR EVERY MONTH-UBER";

**PROC** **SQL**;

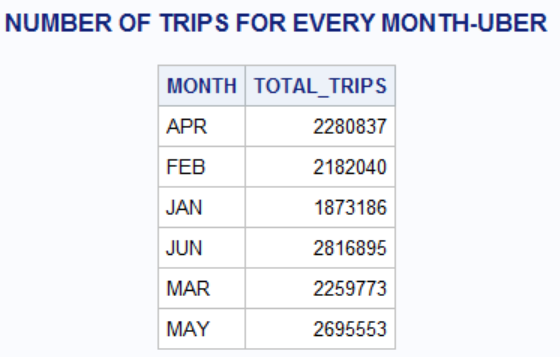
SELECT MONTH AS MONTH,COUNT(&BASE\_NO)AS TOTAL\_TRIPS

FROM UBER\_DATA

GROUP BY **1**

ORDER BY **1** ASC;

**QUIT**;



1. **Total Trips for Every Location**

%LET LOC=locationID;

TITLE"NUMBER OF TRIPS FOR EVERY LOCATION-UBER";

**PROC** **SQL**;

SELECT &LOC AS LOCATION\_ID,COUNT(&BASE\_NO)AS TOTAL\_TRIPS

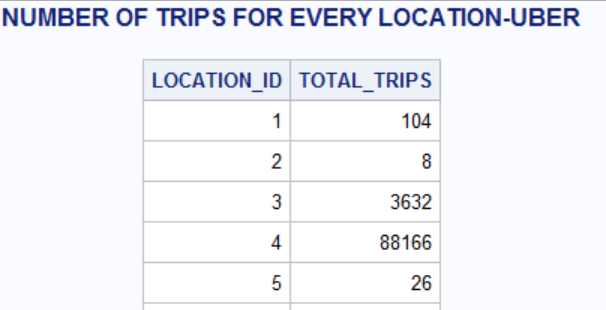
FROM UBER\_DATA

GROUP BY **1**

ORDER BY **1** ASC;

**QUIT**;

**Sample dataset:**



**INFERENCE:**

Merging both the FHV and UBER data sets to compare the number of trips per month:

TITLE "NO OF TRIPS FOR EVERY MONTH-OTHER-FHV & UBER COMPARISON";

**PROC** **SQL**;

SELECT UBER.MONTH,SUM(&TRIPS)AS TOTAL\_TRIPS\_FHV,COUNT(&BASE\_NO)AS TOTAL\_TRIPS\_UBER

FROM OTHER\_FHV\_DATA FHV

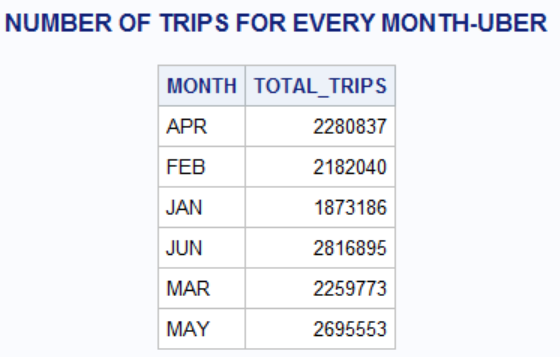
FULL JOIN UBER\_DATA UBER

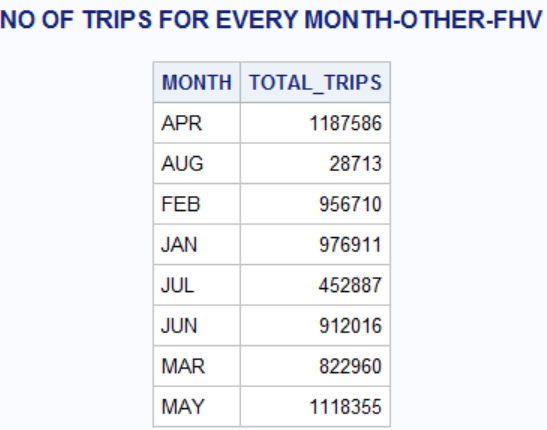
ON FHV.MONTH=UBER.MONTH

GROUP BY **1**

ORDER BY **1**;

**QUIT**;





**INFERENCE:**

* *On Comparing the number of trips taken by Other-FHV services and UBER for the month of Jan to June 2015, we could say that UBER has maximum number of trips for every month compared to other FHV’s.*
* *UBER has influenced many residents of NYC and has been people’s choice by the summary data that we have obtained so far. Business also uses other sets of summary data, for making higher level business decision making, that may help to improve UBER’s business strategy.*