17MDC36 -BUSINESS STATISTICS LAB USING R   
  
  
UBER TRIP PREDICTION AND ANALYSIS

**TEAM MEMBERS:**

**-2033022 NARAYANASWAMY R**

**-2033003 ADHAVA PERUMAL S**

# **PROBLEM STATEMENT**

THE ANALYSIS AND PREDICTION OF TRIPS BRANCHWISE AND IDENTIFICATION OF BRANCH WITH BEST PERFORMANCE HAS NOT BEEN PERFORMED USING RELATION OF TRIPS AND ACTIVE VECHICLES

# OBJECTIVE

* To analyse from the given data set about the profit , trips completed , etc. of UBER transport company
* To predict number of vehicles required in upcoming months.
* To forecast the number of trips that can be achieved with the current number of vehicles available
* To study the relationship between dates and trips

# DATA SET

JAN-FEB TRIP DATA SET TAKEN FROM KAGGLE CONSISITING OF:

1. DATE
2. BASE
3. ACTIVE VEHICLES
4. TRIPS COMPLETED

* SOURCE:

UBER-JAN-FEB-FOIL.csv

[*https://www.kaggle.com/theoddwaffle/uber-data-analysis/data*](https://www.kaggle.com/theoddwaffle/uber-data-analysis/data)

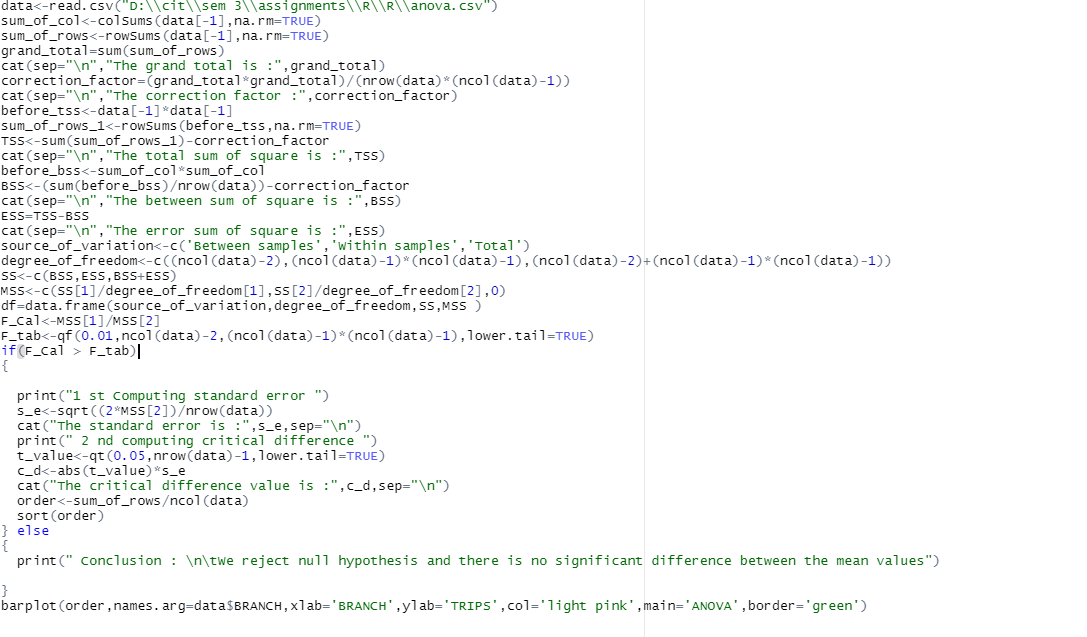
# TECHNIQUES USED

## 1)ONE WAY ANOVA:

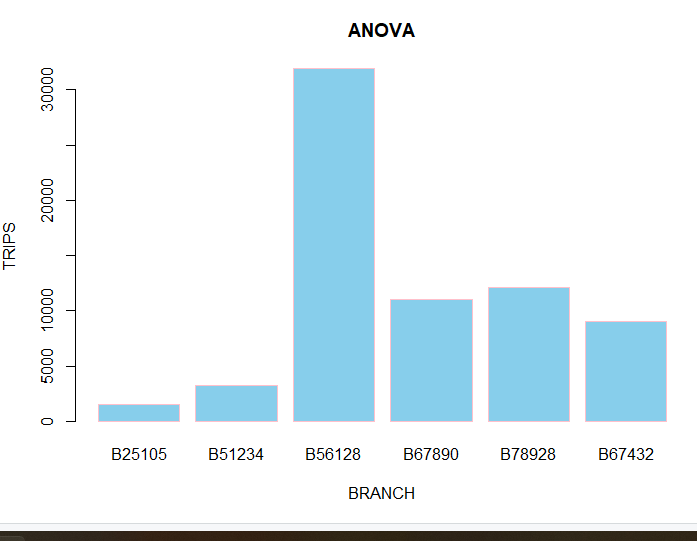
**SOFTWARE USED: R STUDIO**

*IT WAS USED TO CHOOSE BASE WITH HIGHEST PROFIT AND WITH LEAST TRIPS*

**CODE:**

****

### ANALYSIS REPORTS AND INFERENCES:



FROM THE GIVEN DATA,CALCULATING ANOVA,IT IS FOUND THAT THE THERE IS VARIANCE IN NO.OF TRIPS WHEN COMPARING WITH THE BRANCHES.

BY GRAPH,IT IS FOUND THAT BRANCH 3 HAS HIGHEST NUMBER OF TRIPS,I.E PROFIT WHILE BRANCH 1 IS THE LEAST

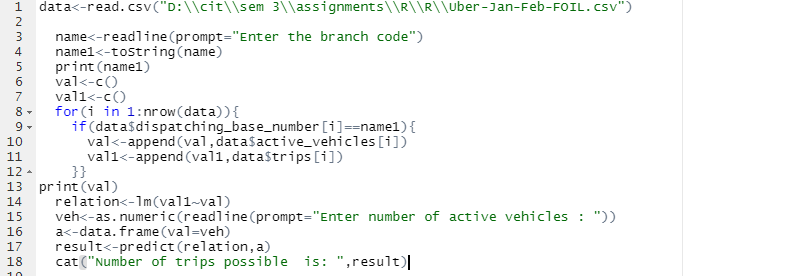
THUS, BRANCH 1 REQUIRES MORE MARKETING AND PROMTIONAL ACTIVITIES

## 2)LINEAR REGRESSION:

**SOFTWARE USED: R STUDIO**

*IT IS USED TO PREDICT TRIPS AND NUMBER OF VEHICLES REQUIRED FOR UPCOMING MONTHS*

**CODE:**



### SAMPLE OUTPUT:

### INFERENCE:

*FROM THE SAMPLE OUTPUT ABOVE,IT IS FOUND THAT,IN THE BRANCH CODE 3,IF 1500 VEHICLES ARE AVAILABLE, IT IS EXPECTED THAT 5724 TRIPS CAN BE COMPLETED.*

*BY REPEATED RUN,WE CAN IDENTIFY THE MINIMUM NUMBER OF VEHICLES REQUIRED AND WILL BE USEFUL FOR MANAGING DAY TO DAY VEHICLE COUNT NEED.*

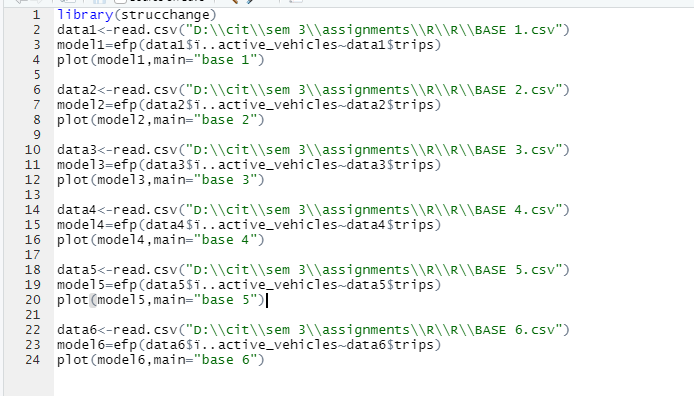
## 3)CUSUM TEST:

**PACKAGE USED: STRUCCHANGE**

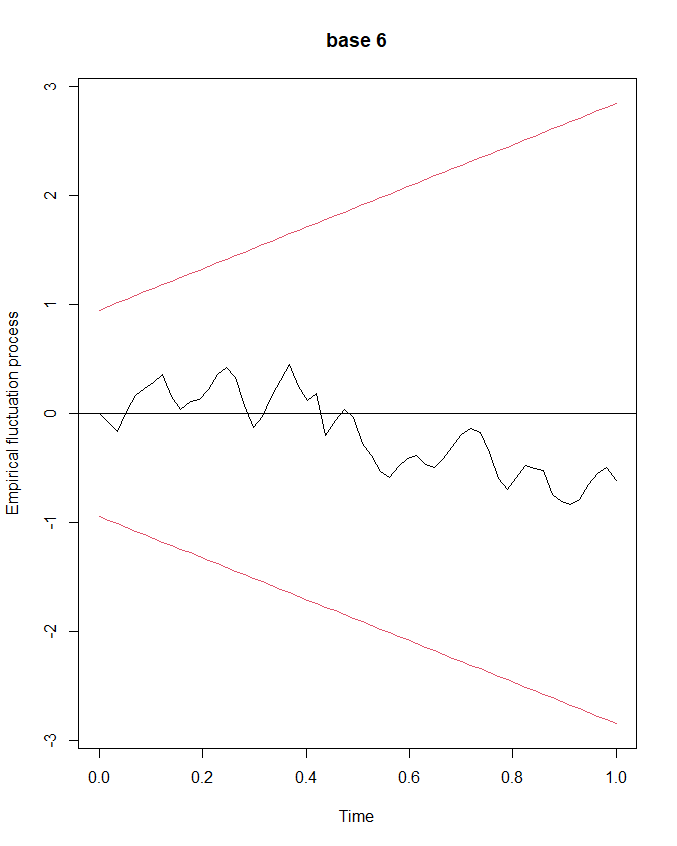
**SOFTWARE USED:R STUDIO**

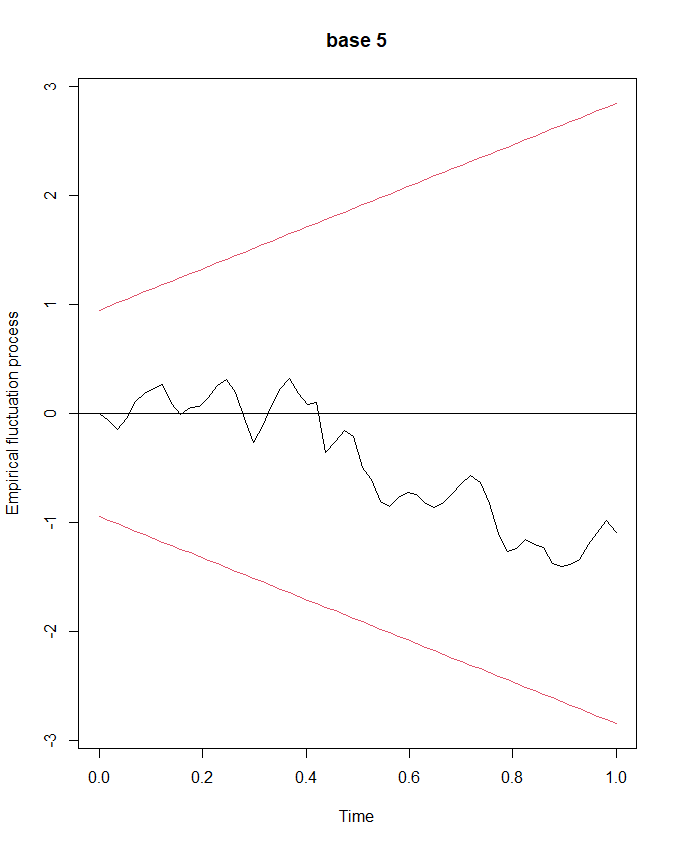
*TO STUDY THE VARIANCE/STABILITY OF TRIPS*

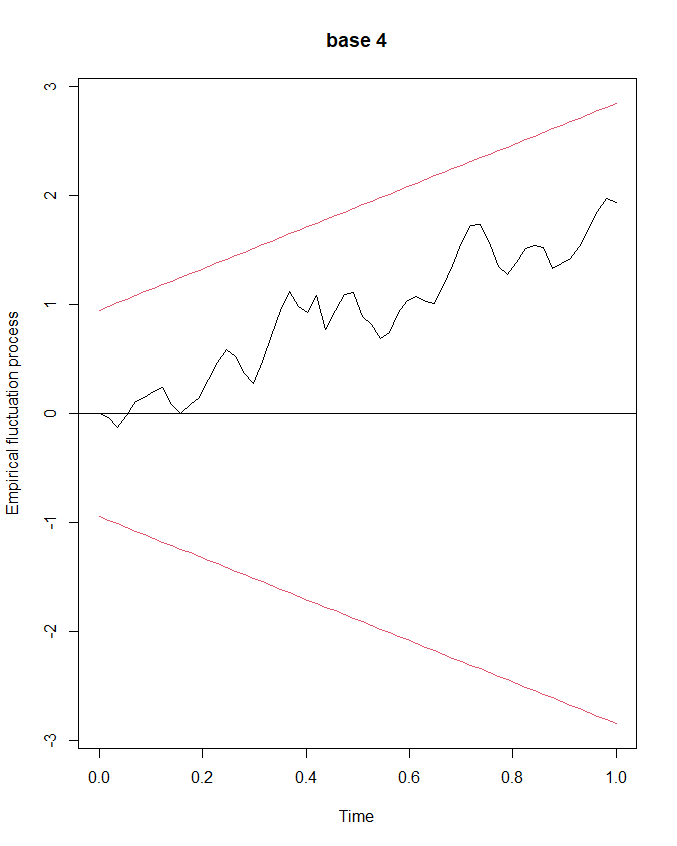
**CODE:**

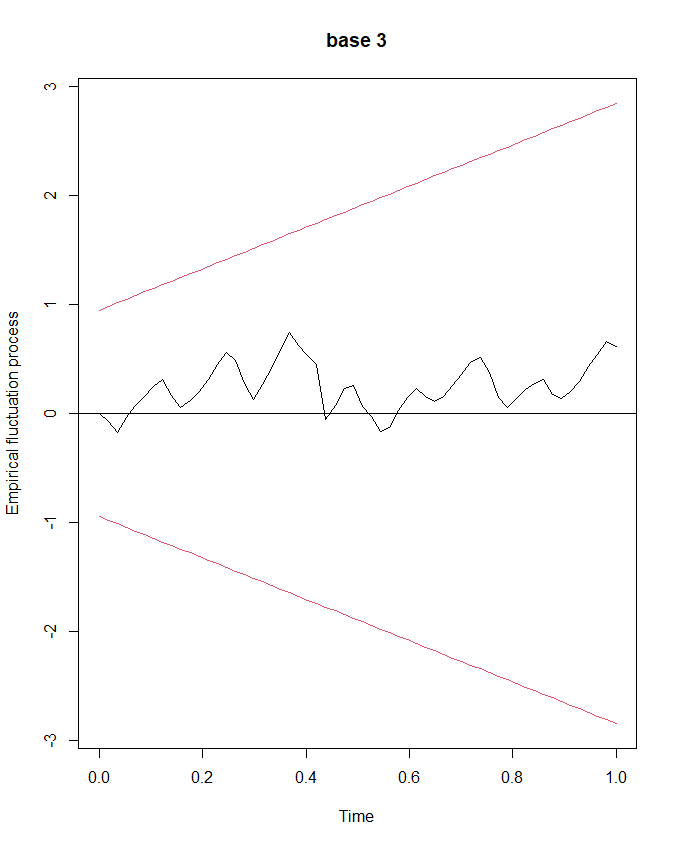


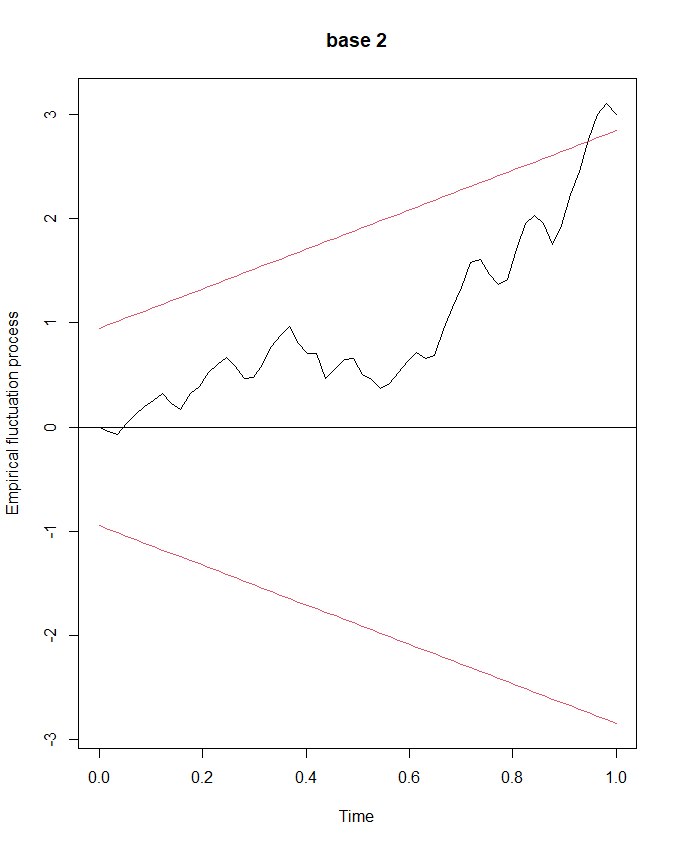
### ANALYSIS AND INFERENCE:

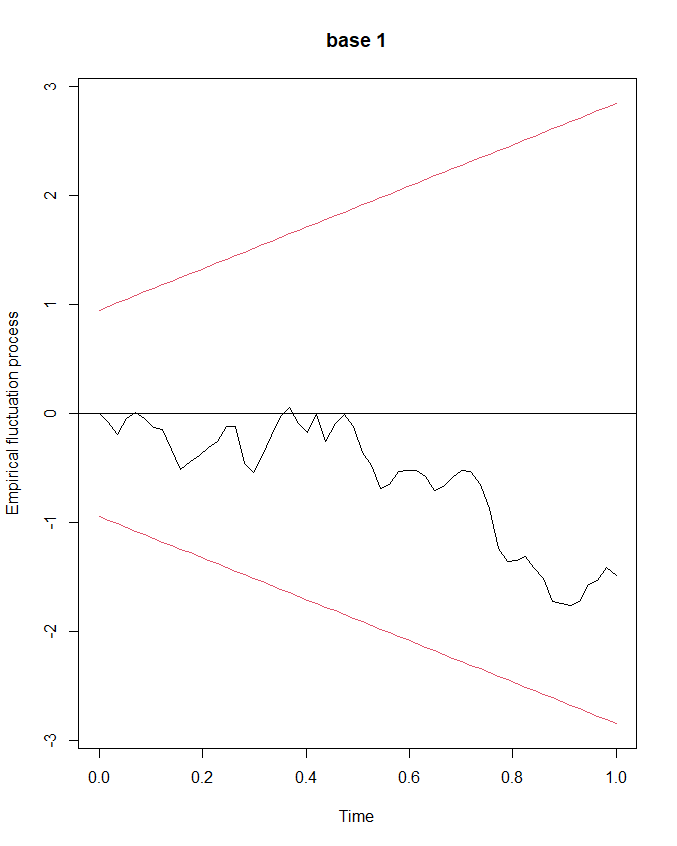












* *EXCEPT FOR BRANCH 2,ALL OTHER BRANCHES HAVE NO MUCH VARIANCE,I.E,STABLE* IN *NUMBER OF TRIPS*
* *BRANCH 2 SHOWS A LOT OF VARIATION IN TRIPS AT MONTH END,THUS,THE VEHICLES MUST ALSO INCREASE AT MONTH END TO AVOID DELAYS IN WAIT TIME*
* *THE REASONS FOR RISE IN TRIPS MUST ALSO BE IDENTIFIED BY LOCAL ANALYSING TEAM*

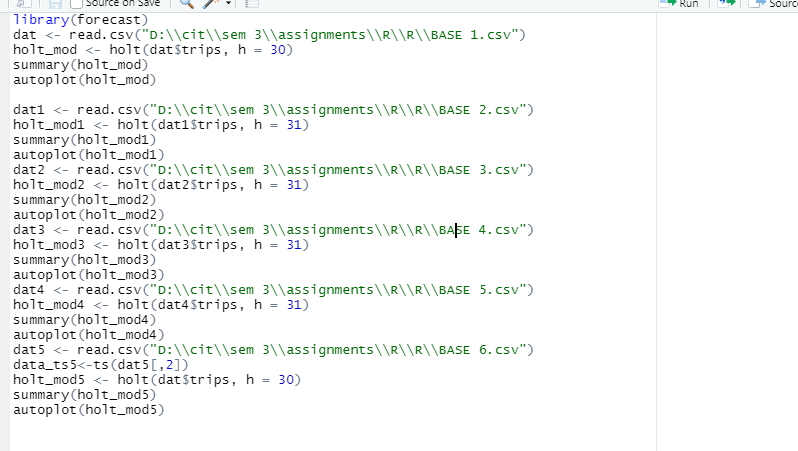
## 4) TIME SERIES FORECASTING:

**LIBRARY USED:FORECAST**

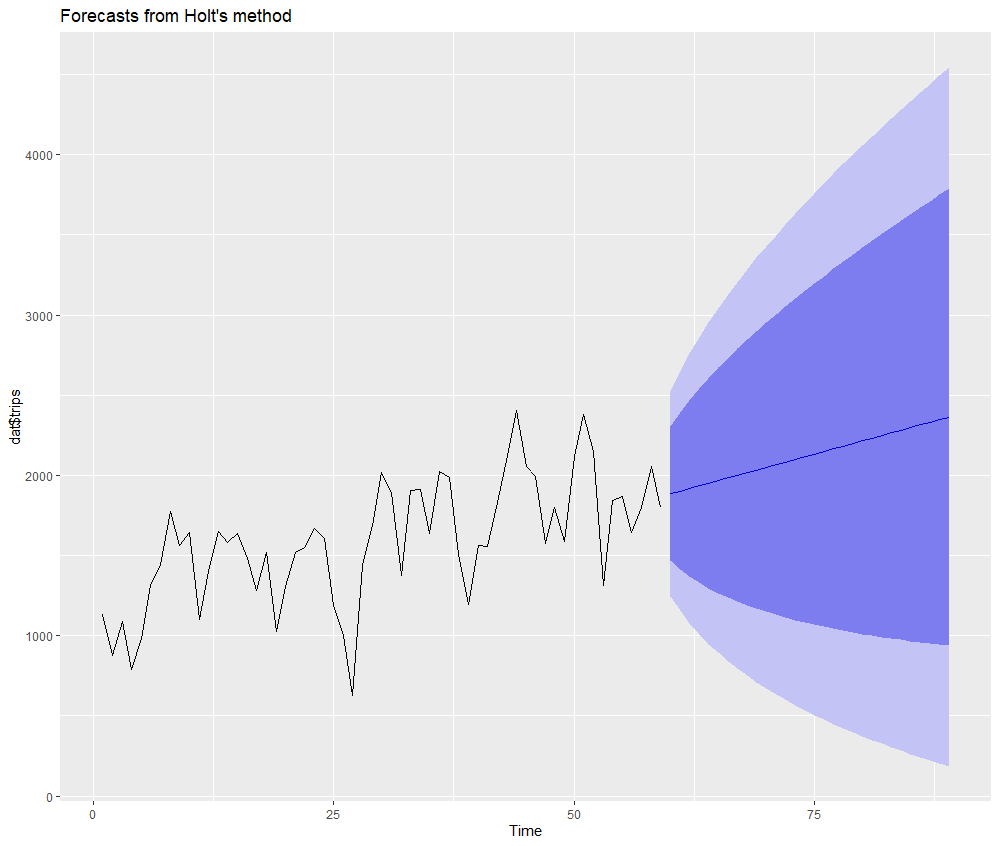
**SOFTWARE USED:R STUDIO**

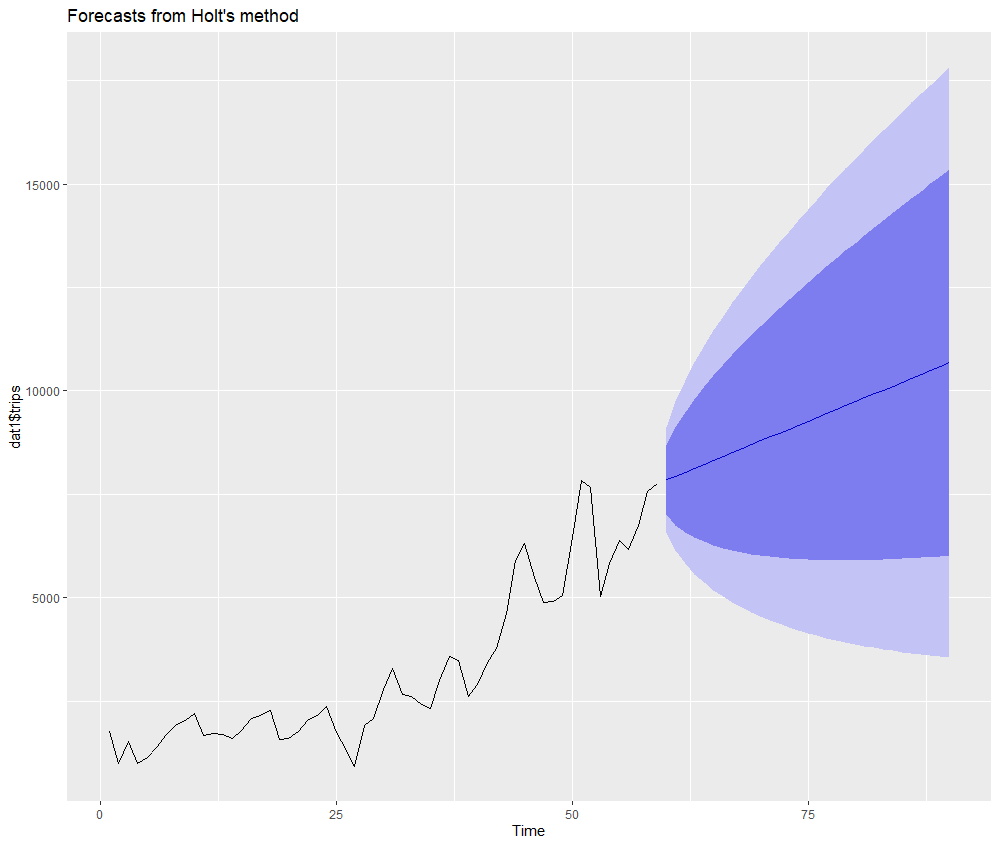
*TO FORECAST THE TRIPS FOR THE NEXT 30 DAYS*

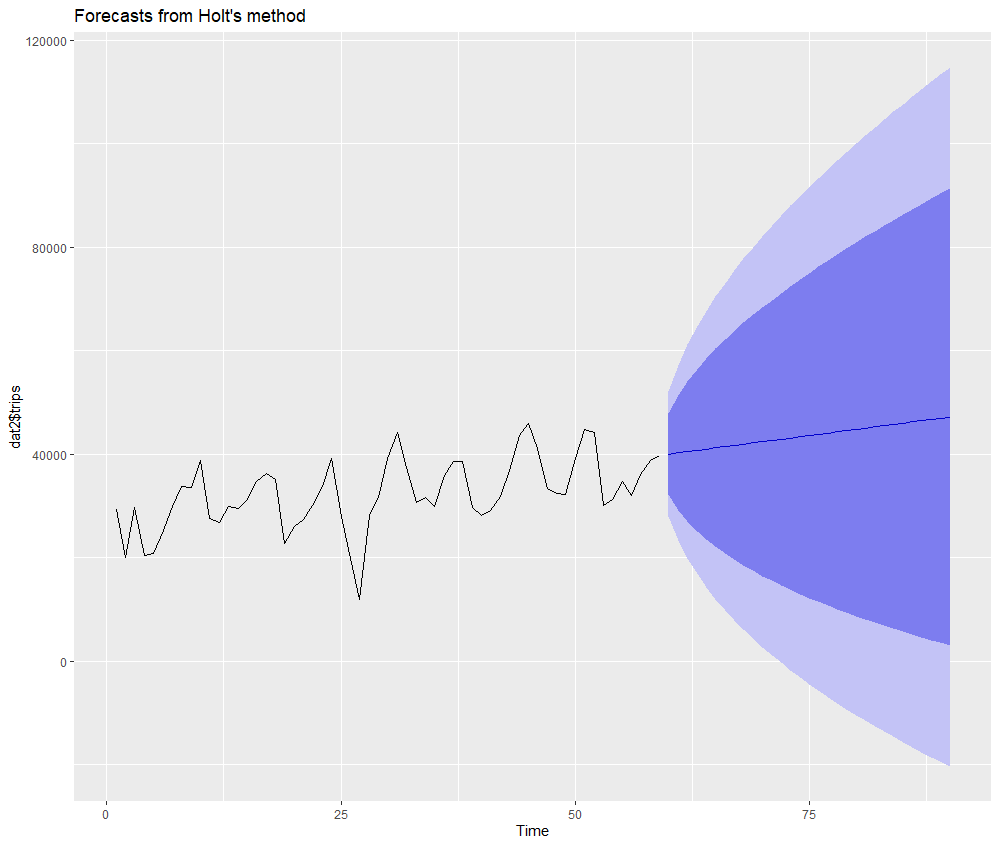
**CODE:**

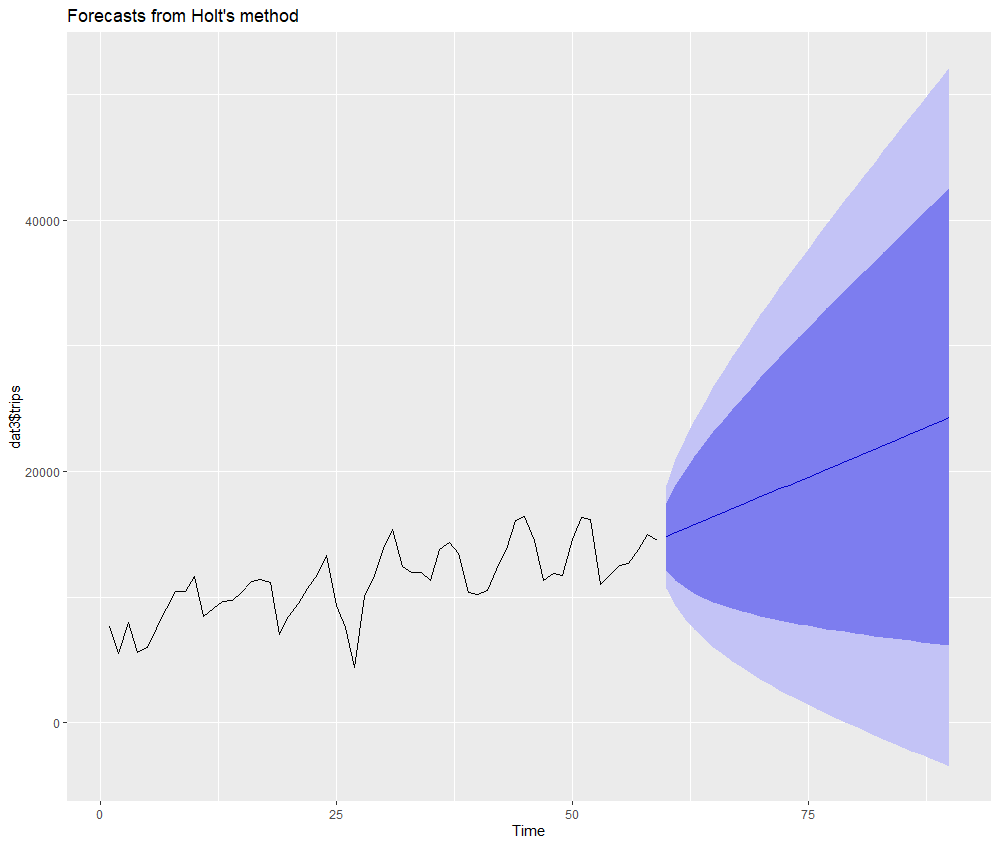
****

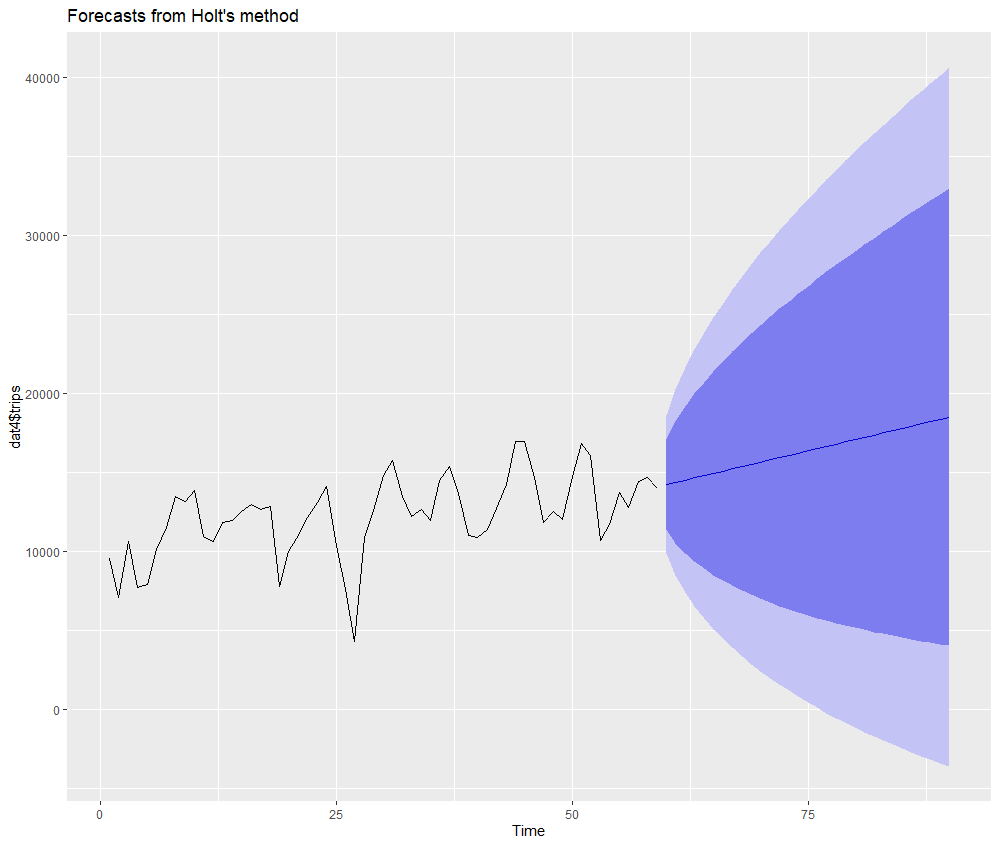
### ANALYSIS AND INFERENCE:

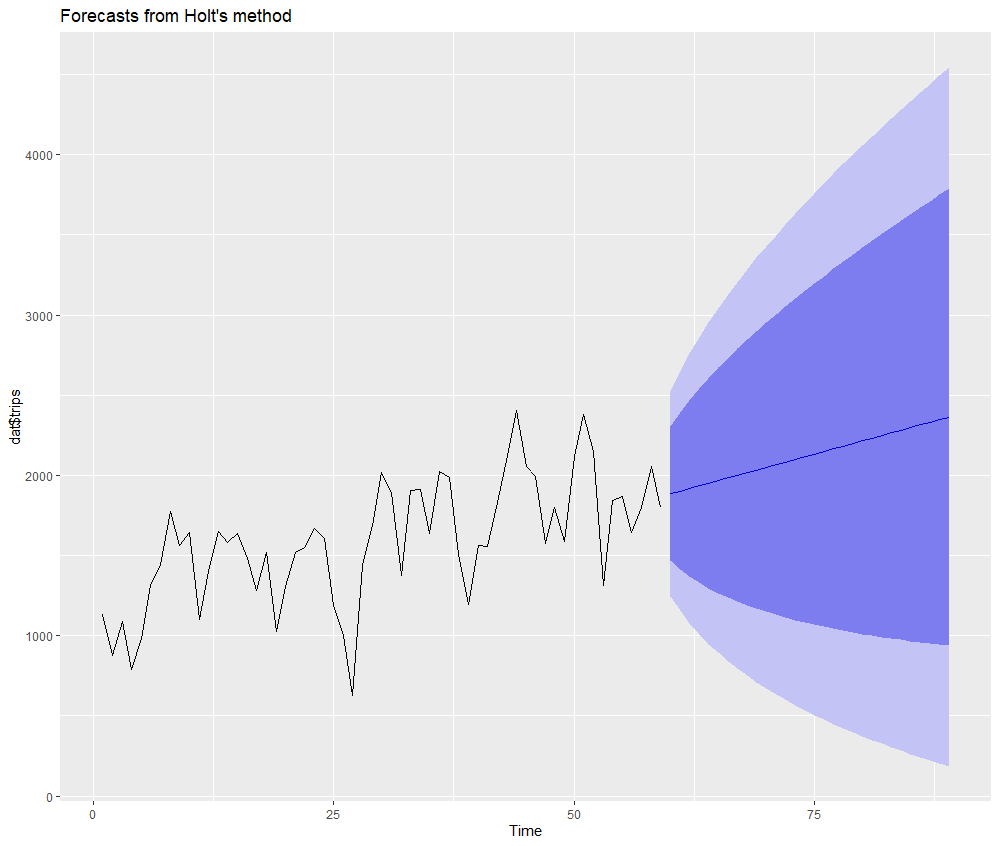












1. *IT IS FOUND THAT FOR THE NEXT 30 DAYS,THE TRIPS WOULD DEFINITELY INCREASE STEADILY*
2. *THUS,THE COMPANY IS NOT EXPECTED TO BE IN LOSS FOR NEXT 30 DAYS*

## 5)CORRELATION:

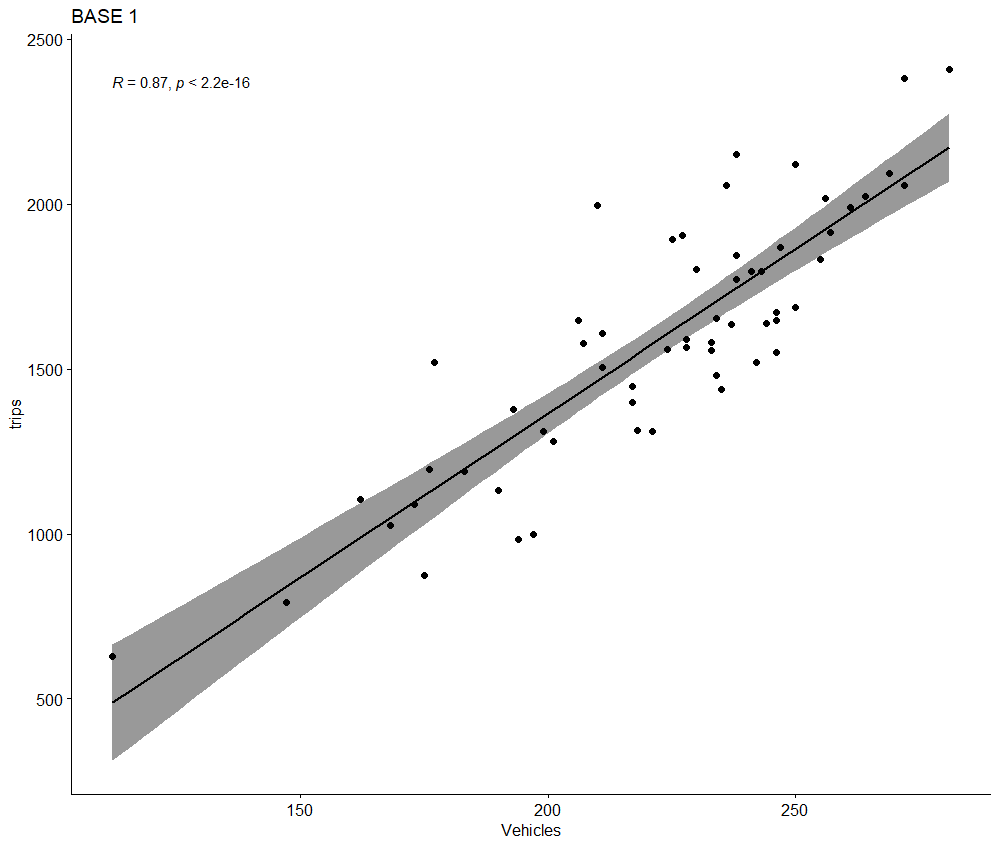
**LIBRARY USED:GGPLOT2**

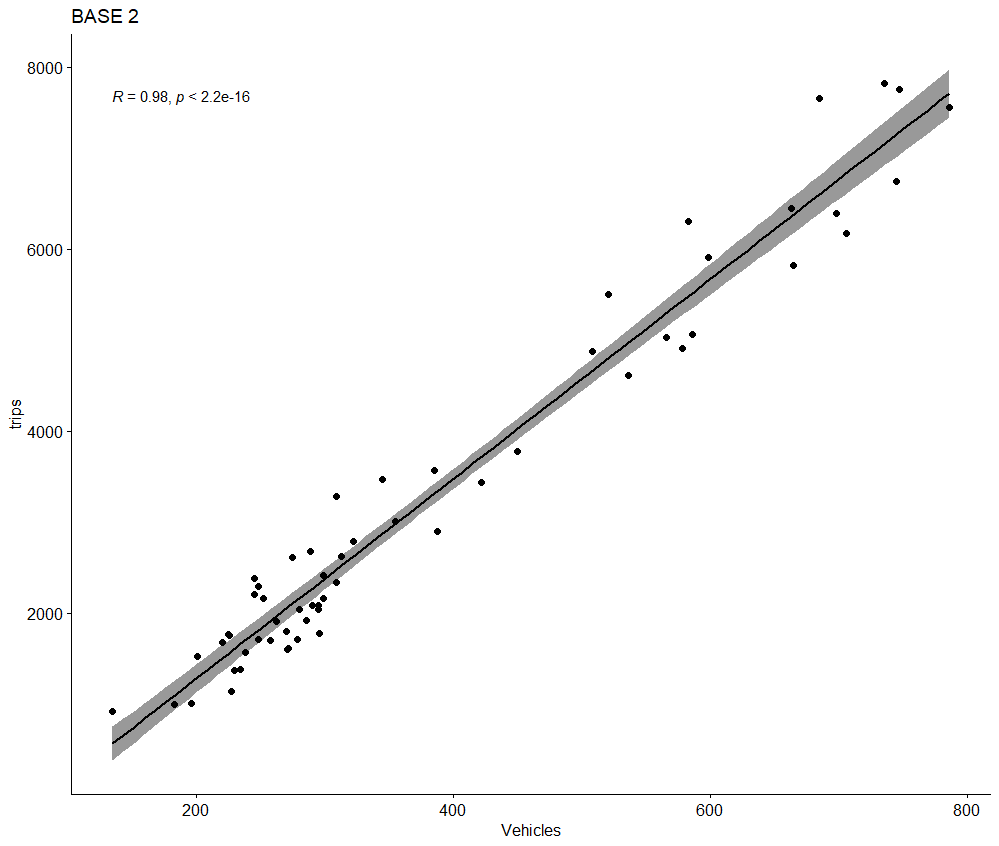
**SOFTWARE USED: R STUDIO**

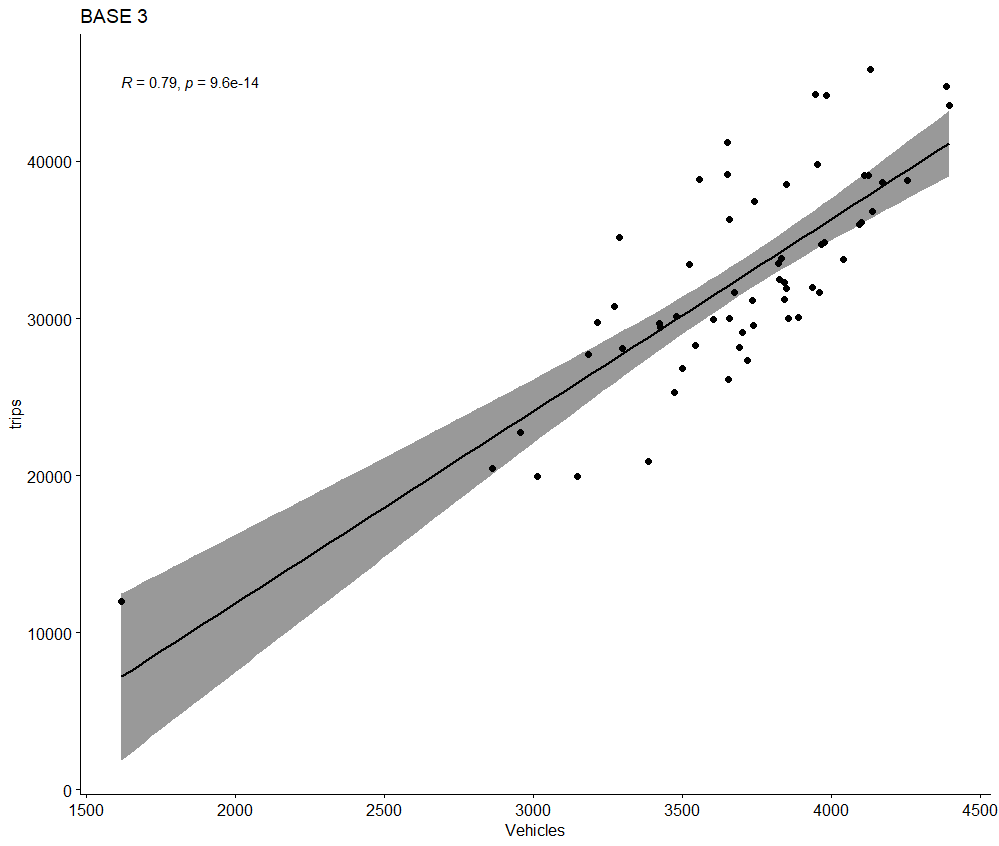
*TO STUDY THE RELATIONSHIP BETWEEN TRIPS AND VEHICLES*

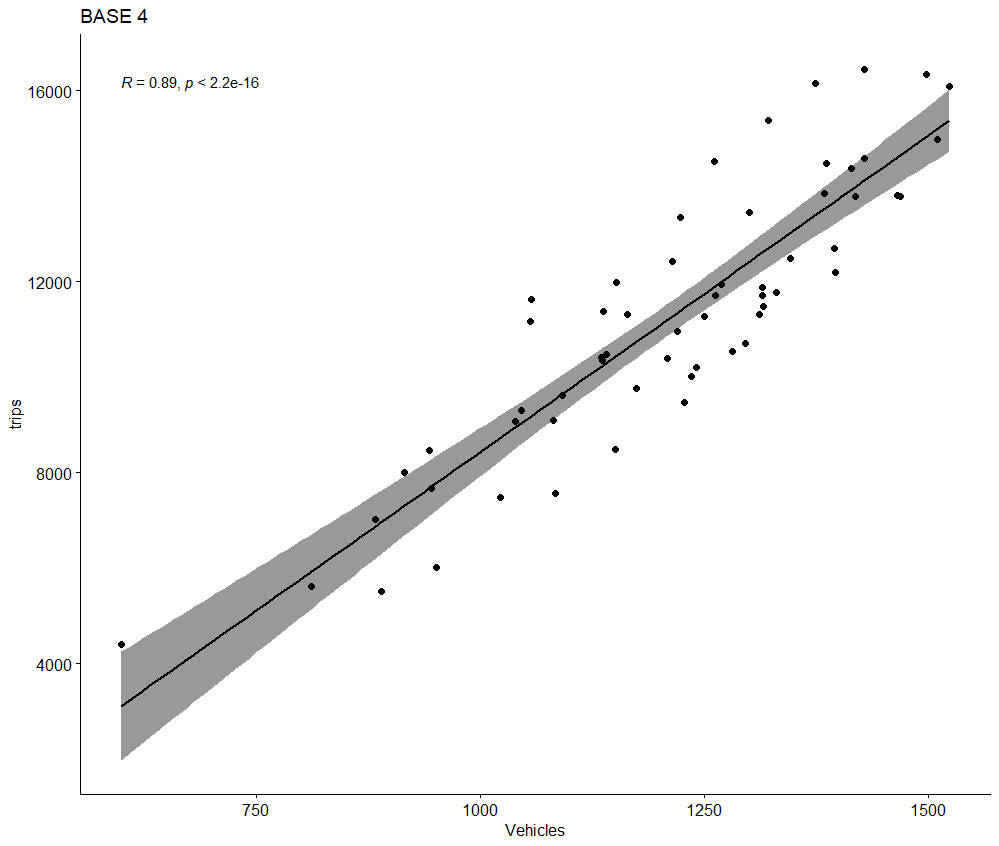
**CODE:  
**

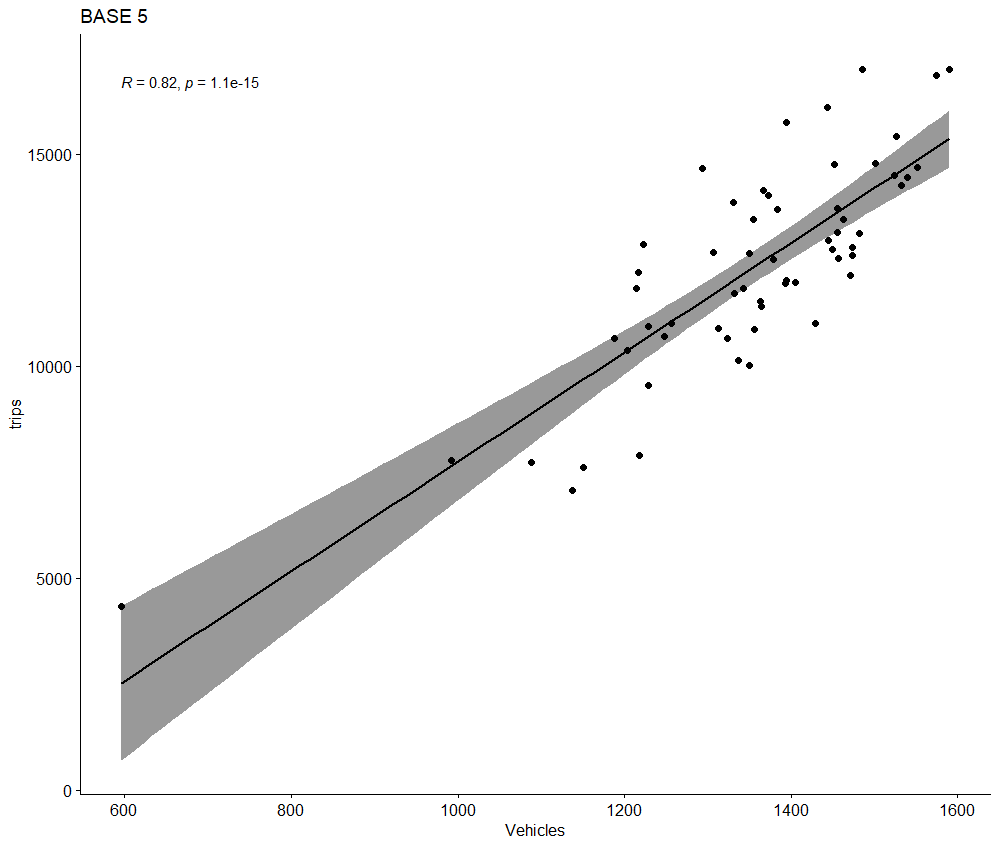
### ANLAYSIS AND INFERECE:

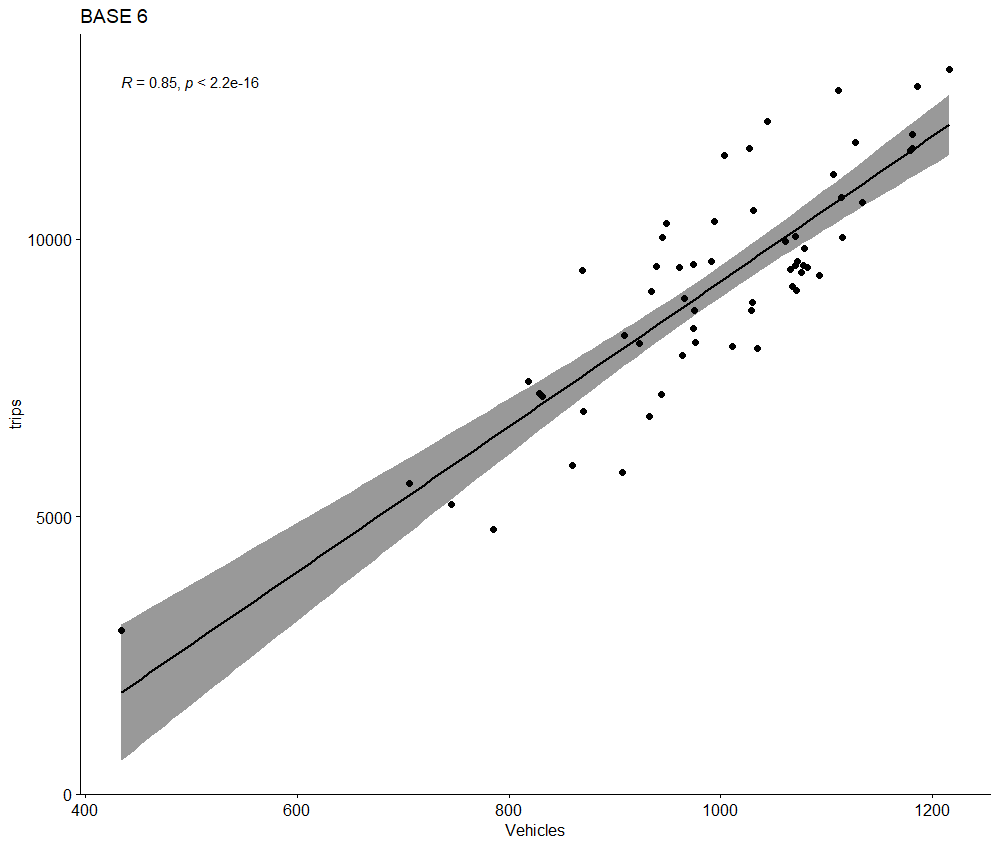












* *ALL THE BRANCHES’ TRIPS AND VEHICLES ARE POSITIVELY CORRELATED*
* *THIS MEANS THAT IF VEHICLES ARE INCREASED, TRIPS WILL ALSO INCREASE AND TRIPS WILL DECREASE WHEN VEHICLES COUNT DECREASE*
* *THUS,INCREASING VEHICLES WOULD BOOST THE INCOME OF COMPANY*

# CONCLUSION

1. BRANCH 1 REQUIRES MORE MARKETING AND PROMTIONAL ACTIVITIES
2. THE VEHICLES MUST ALSO INCREASE AT MONTH END TO AVOID DELAYS IN WAIT TIME
3. THE REASONS FOR RISE IN TRIPS MUST ALSO BE IDENTIFIED BY LOCAL ANALYSING TEAM
4. INCREASING VEHICLES WOULD BOOST THE INCOME OF COMPANY
5. THE COMPANY IS NOT EXPECTED TO BE IN LOSS FOR NEXT 30 DAYS

**\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\*\***