

# Daily Trips: How often do Americans Leave Home?

By: Trevor McCormick, Dr. Robert Kelley

Bellarmino University: Data Science Program



## Abstract

- Most people in America leave their houses on a regular basis, but some do so at a much higher or lower rate than others. The Department of Transportation has recorded data on how often people leave their homes daily in each state.
- The goal of this project was to see if machine learning can be used to predict how often people will leave their homes in the future on a given date.
- In this project I will be using various applications such as Python (with packages), SQLite, and Power BI to do my analysis, visualization, and testing.



## Applications and Materials

**Data:** "Trips by Distance" by data.gov

- Department of Transportation

**Clean-Up/Analysis:** Python

- Pandas, Matplotlib, NumPy

**Storage:** SQLite

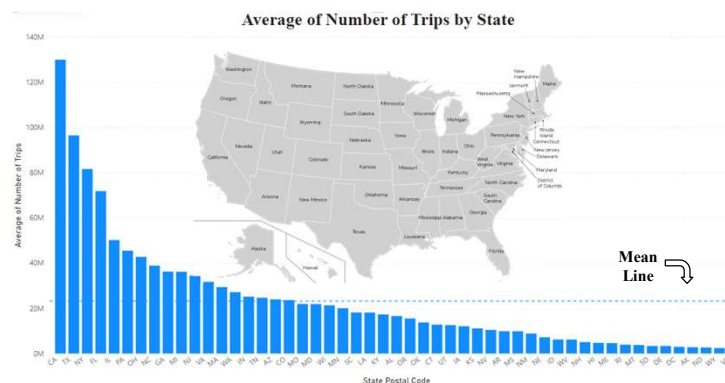
**Visualization:** Power BI

**Machine Learning Tools:** Random Forest and/or Decision Tree

## Objectives

- Import the Trips Data from Data.gov into Python and make a Pandas data frame with the data inside
- Store the Pandas data frame in the SQLite application
- Pull the data from SQLite and import it back into Python for cleaning, exploratory analysis, and basic visualization
- Use Power BI for data visualization by making graphs and charts that accurately depict the Trips Data
- Run machine learning techniques to experiment with and test the data with the goal of predictive analysis to predict future trips

## Results



- After cleaning and analysis, the trips data set ended up with a total of:
  - 20 Columns
  - 45,267 Rows
  - 0 Missing Values
- Each states has exactly 887 values for each column
- After being imported into Power BI, the data was converted into charts in graphs like the one found above.
  - This graph shows the average number of trips taken in each of the 50 states
- Machine Learning techniques were used on the data in Python for the predictive analysis step of this project. This will be discussed in the "Conclusions/Future Work" section.

## Conclusions/Future Work

- Random Forest and Multiple Linear Regression were used as the final predictive models for this dataset
- The Random Forest Model yielded a predictive score of 98%
- The Multiple Linear Regression Model yielded a predictive score of 94%
- Given that the score for the Random Forest Model is almost suspiciously high, I think that the best model overall to use for this dataset is the Linear Regression model
- In the future, I would give myself more time on the modelling section of the project in order to explore all of the potential options more thoroughly

## References

- <https://catalog.data.gov/dataset/trips-by-distance>
- <https://sqlite.org/index.html>
- <https://www.python.org/>
- <https://powerbi.microsoft.com/en-us/>

## Author Contact

- ❖ **Phone:** 502-902-7455
- ❖ **School Email:** tmccormick2@bellarmine.edu
- ❖ **Personal Email:** trevormccormick01@gmail.com

## Acknowledgements

Thanks to Bellarmine University, Dr. Kelley, Dr. Fenton, and all the other Bellarmine professors and faculty who have guided me through my academic career!