**PROJECT REPORT: DATA TO DOLLARS**

**What Are We Studying?**

In this project, we are diving into a dataset that gives us a closer look at NYU Taxi rides, including key details like how long the trip was, how much it cost, how many passengers were in the car, and whether they paid by cash or card. But the most important thing we are focusing on is the **tip amount** — the extra money passengers give their drivers.

We will look closely at Independent Variables like **fare Amount**, **trip distance**, **Passenger count, pick-up Date & Time, Drop Off Date & Time, and Congestion Surcharge** to see how each influences the Tip.

**Why Does It Matter?**

We are aiming to figure out what makes passengers more likely to tip generously. For example, by focusing on things like suggesting card payments or choosing longer routes, drivers could potentially earn more in tips.

On a bigger level, understanding tipping helps us see how people make decisions in various situations. Whether it is the price of the ride, how long the trip takes, or how easy it is to pay by card, these factors give us a better understanding of why people choose to tip the way they do.

**How Do We Study This?**

We are using a dataset that covers lots of details about NYU Taxi trips, capturing all the key information about each trip. It includes things like when and where the ride started and ended, the peak hours affecting tip, how far the taxi traveled, and how much the fare was. On top of that, we can see whether passengers left tips, with regards to the duration of the ride, and even how they paid — whether with cash or card.

We are using a dataset that captures a wide range of details about NYC taxi trips, including Variables such as:

**Fare Amount**: The base fare for the trip.

**Trip Distance**: The total distance of the trip in miles or kilometers.

**Passenger Count**: The number of passengers in the taxi.

**Pickup date & time:** The date and time when the passenger was picked up.

**Drop-off date & time:** The date and time when the passenger was dropped off.

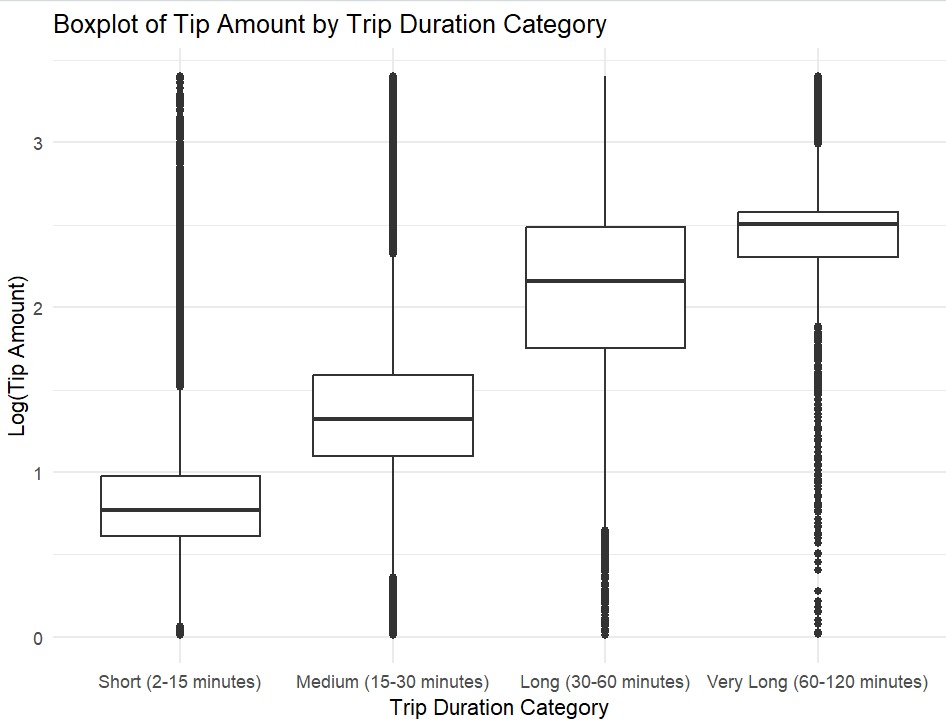
**Congestion Surcharge**:  An additional charge for trips taken during high traffic congestion times.

**Tip amount** (the outcome) and figuring out what affects it.

**What Did We Find Through Data Visualization?**

Here is what our analysis showed:

* **Does longer Duration lead to better tip amount?**

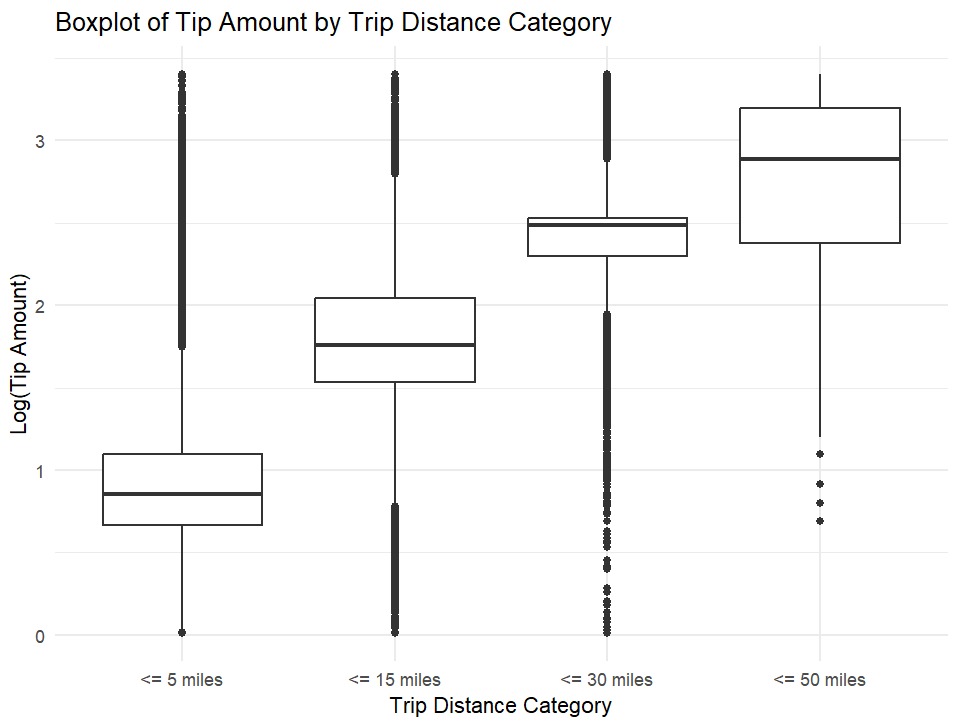


There is a clear upward trend in median tip amounts as trip duration increases, suggesting that longer trips generally result in higher tips.

The variability in tip amounts also increases with trip duration, especially for trips longer than 30 minutes, indicating more diverse tipping behavior for extended rides.

The presence of many high outliers in the "Very Long" category suggests that customers are willing to tip generously for very long trips, even though the range of tip amounts is wide.

* **Are longer trips rewarded with higher tips?**

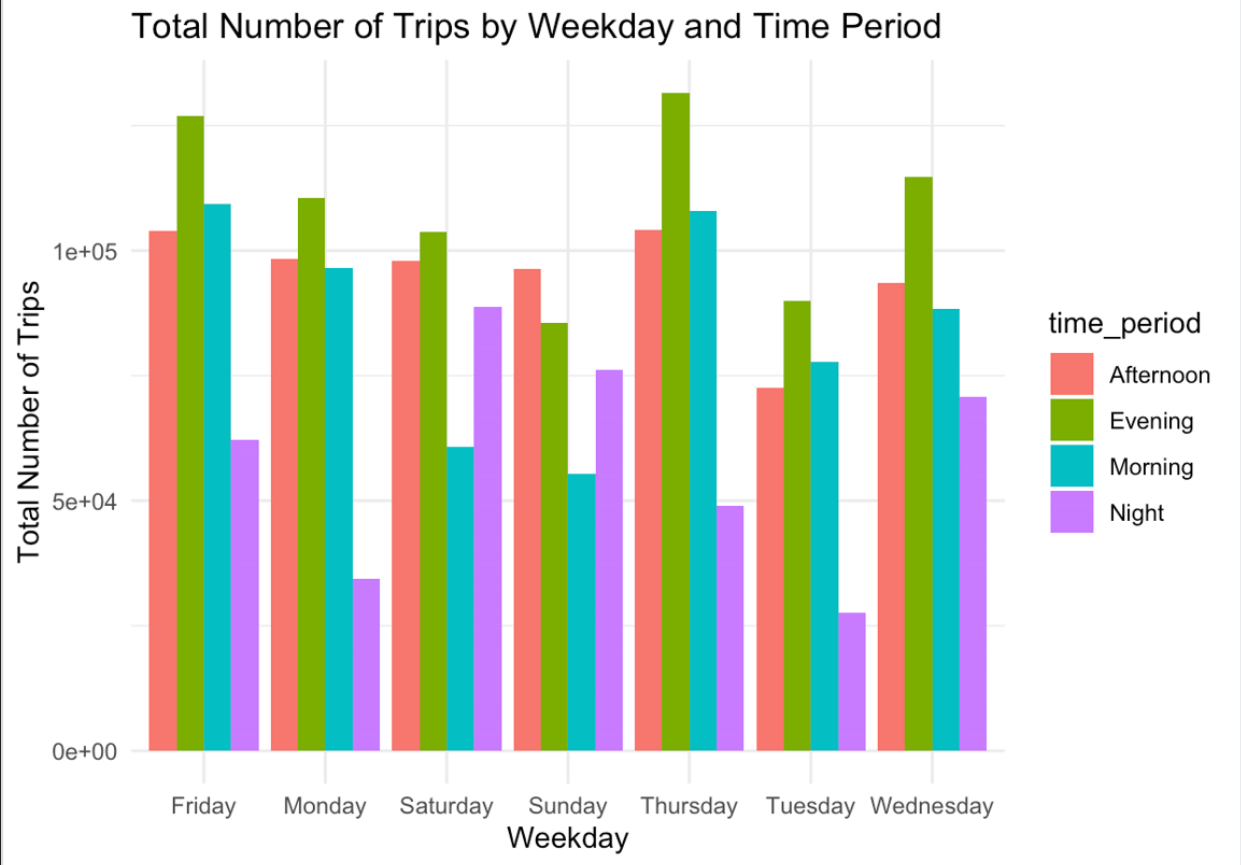


There is a clear positive relationship between trip distance and tip amount; longer trips generally result in higher tips.

1. The number of outliers (very high tips) increases with the trip distance, suggesting that customers may tip generously for extended journeys.
2. Short trips have more clustered tips around lower values, possibly indicating standard tipping behavior for routine commutes.

* Do passengers who pay with a card tip more than those who pay with cash?
* Does the number of people in the taxi affect how much the driver is tipped?
* **For the rush hour analysis:**

Morning and Evening are considered rush hours due to people commuting to and from work, which explains why the number of trips is significantly higher during these periods on weekdays. However, on Saturday and Sunday, there is a noticeable drop in morning trips, as people tend to have more relaxed schedules on weekends, resulting in less morning traffic. In contrast, the afternoon and evening periods maintain higher trip counts on weekends, reflecting leisure activities and social outings.



* **For the tipping analysis:**

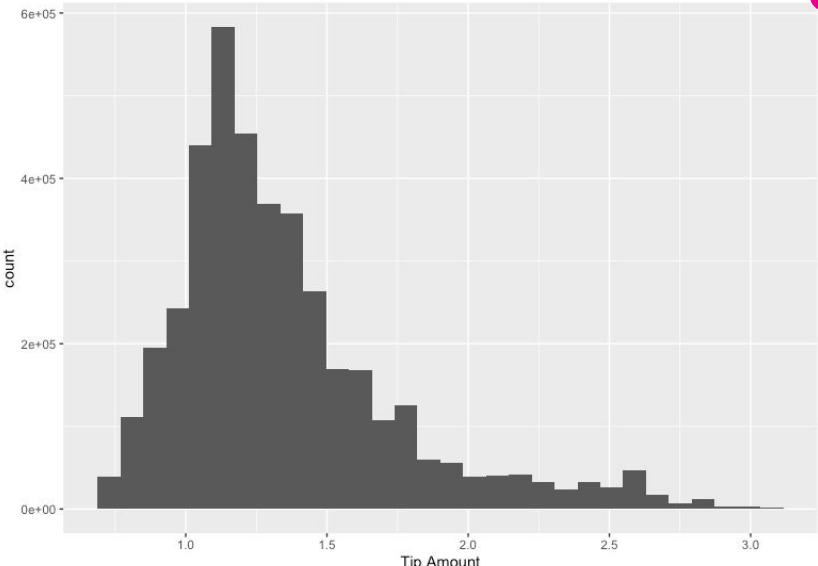
Most tips are small: The majority of tips are concentrated around 1 unit, indicating that small tips are the norm.

Right-skewed distribution: The graph shows a clear right-skew, with fewer tips at higher amounts, indicating that larger tips are rare.

Sharp drop after 1.5 units: There is a noticeable decline in tip frequency after 1.5 units, suggesting that few customers tip more than this amount.

Large tips are uncommon: Very few tips exceed 2 units, reinforcing that generous tipping is relatively rare in this dataset.

High peak at 1 unit: The highest number of tips is around 1 unit, with over 600,000 occurrences, highlighting this as a common tipping behavior.



**What Does It Mean?**

* **Positive relationship between trip distance and tips**: Longer trips generally lead to higher tips, as passengers may feel drivers deserve more compensation for extended journeys.
* **Outliers (very high tips) increase on longer trips**: These unusually high tips become more frequent for extended rides, suggesting passengers might tip generously for long journeys, possibly out of appreciation for the extra service.
* **Short trips typically have smaller, standard tips**: Tips for shorter rides tend to be lower and more consistent, likely because these trips are routine, like commutes or errands.
* **Tips are directly linked to fare amounts**: There's a direct connection between the fare amount and the tip, with higher fares often resulting in higher tips.
* **Tips vary depending on the day and rush hour patterns**: Tipping behavior also changes based on the day of the week. Weekday rush hours (morning and evening) see more trips and, consequently, more tips, especially in the evenings when passengers might be more generous after work.
* **Increasing median tip amounts for longer trips**: As trip durations increase, so do the median tips, indicating passengers recognize the value of longer trips and adjust their tipping accordingly.
* **Greater variability in tips for extended trips**: For trips lasting over 30 minutes, there’s more variety in tipping behavior. Some passengers leave generous tips, while others may not follow a similar pattern, leading to a broader range in tips.
* **Very long trips often result in higher tips**: While the range of tips varies widely for very long rides, a notable number of passengers tend to leave substantial tips, rewarding drivers for the extended service.
* **A model analyzing significant factors**: A detailed model was created to explore how key factors like trip duration, distance, and time of day influence tipping behavior, highlighting independent variables that strongly impact the tip amount.