Daniel Rojas, Ryan Huang, Riley Menter, Tyler Pak

Professor: Dr. Xu Cao

March 13, 2025

Optimizing UberEats Delivery Times

BUS 127 - Introduction to Quality Control

Section: 24 Friday 3 PM

Professor Xu Cao

Introduction:

In this project, we will analyze the efficiency of Uber Eats delivery times in Southern California using simulated data. More specifically, we focus on two key factors, delivery time and its corresponding distance. To do this, we will look at how many minutes it takes to travel one mile on average for each delivery. From the personal experiences of our group members, Uber Eats delivery times can fluctuate significantly, even for deliveries with similar distances. Moreover, several posts and articles online discuss the phenomenon of slow delivery times. This inconsistency can impact customer satisfaction. It can also affect food quality and overall service quality. By analyzing variables and considering external factors, we could identify the patterns and inefficiencies in the delivery process for Uber Eats.

We decided to examine Uber Eats' delivery times because of its growing popularity within urban societies. "The last-mile challenge is further exacerbated by the desire for ever faster delivery, to compete with the instant gratification offered by brick-and-mortar stores" (Reyes, Savelsbergh, & Toriello, 2017). One of these popular urban societies is Southern California, which just so happens to be the geographical constraint of our analysis. Delivery drivers deal with many environmental factors such as weather, traffic, and other conditions. Using statistical methods, we aim to uncover insights into how these factors impact delivery times. Our findings can guide Uber Technologies in optimizing logistics, enhancing customer experience, and ensuring more efficient deliveries by reducing delivery times.

Methodology & Data

To analyze Uber Eats delivery efficiency, we simulated our data using R Studio and carefully chose set distributions and probabilities for the variables. We created a 50-observation dataset containing the time of day, delivery distance, delivery time, weather conditions, and traffic status as our variables.

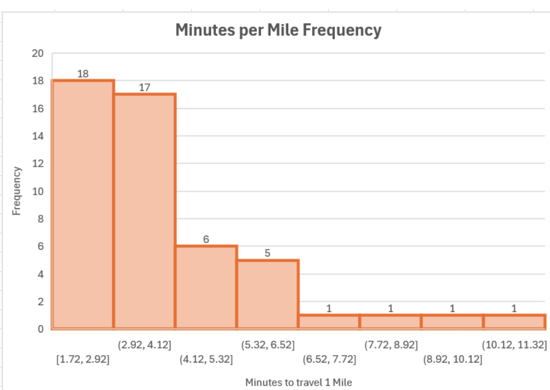
```r
# Simulate first 1/4 delivery data (1-3 miles)
delivery_data1 <- data.frame(
  Observations = 1:n1,
  Time_of_Day = generate_categorical(n1, time_of_day_levels, probs = c(0.05, 0.5, 0.3, 0.15)),
  Distance = runif(n1, min = 1, max = 3), # in miles
  Time = runif(n1, min = 5, max = 15), # in minutes
  Weather = generate_categorical(n1, weather_levels, probs = c(0.6, 0.2, 0.15, 0.05)), # Weather mostly clear
  Traffic = generate_categorical(n1, traffic_levels, probs = c(0.3, 0.7))  # Usually bad traffic
)

# Simulate second 1/4 delivery data (4-6 miles)
delivery_data2 <- data.frame(
  Observations = (n1 + 1):(n1 + n2),
  Time_of_Day = generate_categorical(n2, time_of_day_levels, probs = c(0.05, 0.25, 0.35, 0.35)),
  Distance = runif(n2, min = 4, max = 6),
  Time = runif(n2, min = 10, max = 20),
  Weather = generate_categorical(n2, weather_levels, probs = c(0.6, 0.2, 0.15, 0.05)),
  Traffic = generate_categorical(n2, traffic_levels, probs = c(0.4, 0.6))
)
```

To do this, we created 4 subsimulations of data (the first two are shown above), each corresponding to the intervals of delivery distances: [1:3], [4:6], [6:8], and [8:12] miles. The first two groups (1-6 mile deliveries) generate 40 observations, as most deliveries fall within these distances. The last two groups (6-12 mile deliveries) generate only 10 observations, as it's rare to have deliveries of this distance. The time of day probabilities chosen for each sub-simulation are based on how likely you are to receive a delivery of that distance for each time of day. Shorter distance deliveries are more common in the morning and afternoon, while longer distance deliveries are more common in the evening and night. For the [1:3], [4:6], [6:8], and [8:12] mile sub-simulations, we have the range of possible delivery times set to [5:15], [10:20], [15:25], and [17.5:35] minutes, respectively. These ranges replicate real-life delivery scenarios where shorter-distance deliveries sometimes have longer delivery times. Since Southern California has relatively consistent weather patterns year-round, the probabilities for the 4 weather conditions, Clear, Cloudy, Rainy, and Foggy, are 0.6, 0.2, 0.15, and 0.05, respectively. After combining the sub-simulations into a single dataset, we created a CSV file using an R command, storing it in our local drives. We then opened the file in Excel (Data shown below) to perform exploratory data analysis and statistical analysis.

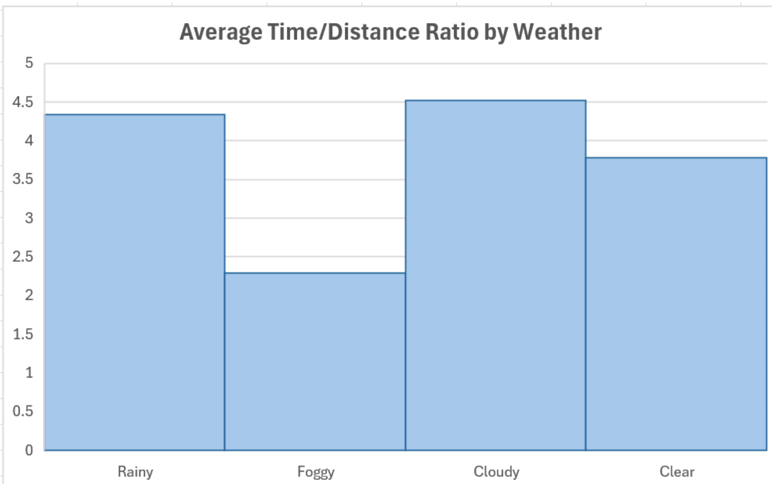
	A	B	C	D	E	F	G	H
1	**DATA:**							
2	Observations	Time_of_Day	Distance	Time	**Time/Distance**	**MR**	Weather	Traffic
3	1	Afternoon	2.829612087	14.04031387	4.96192179		Cloudy	Bad
4	2	Afternoon	2.874150827	6.387101677	2.22225696	2.739664831	Foggy	Bad
5	3	Afternoon	1.57227907	14.88891729	9.469640331	7.247383372	Cloudy	Bad
6	4	Morning	2.660895252	14.46668233	5.436772573	4.032867759	Clear	Bad
7	5	Afternoon	2.283491038	5.824375581	2.550645255	2.886127317	Rainy	Normal
8	6	Morning	2.038191898	10.14211784	4.976036777	2.425391521	Clear	Bad
9	7	Night	2.473176629	8.902034671	3.599433443	1.376603333	Clear	Bad
10	8	Evening	1.269333194	14.05738131	11.07461884	7.475185396	Rainy	Bad
11	9	Morning	2.313984581	9.469696281	4.09237657	6.98224227	Cloudy	Bad
12	10	Evening	2.410129568	13.3600426	5.543288119	1.450911549	Clear	Bad

To analyze the data, we used several visualizations to understand the relationships between the variables. We made a histogram (below on the left) displaying the distribution of time/distance values for deliveries. The distribution (using bin sizes of 1.2) has a right skew, which is a good sign in this situation. Most deliveries are completed, taking 1.72 to 4.12 minutes to travel one mile. The outliers causing the skew are the few deliveries completed, taking 6.52 to 11.32 minutes to travel one mile.



The box plots (above to the right) display the five-number summaries for delivery distances in the afternoon, morning, evening, and night. The 4 plots show that longer-distance deliveries tend to happen more in the evening and night, followed by the afternoon. Using the AVERAGEIF()

function in Excel, a bar plot is generated (below) that examines how the time/distance value is influenced by weather. Because only two of our observations have foggy conditions, measuring delivery performance for this condition is difficult. For the other categories, we can see that rainy and cloudy days, on average, have slightly higher time/distance numbers than clear days. This suggests that the weather has a subtle influence on delivery performance.



Weather	Minutes
Rainy	=AVERAGEIF(G3:G52, J3, E3:E52)
Foggy	2.293086
Cloudy	4.519281
Clear	3.78143

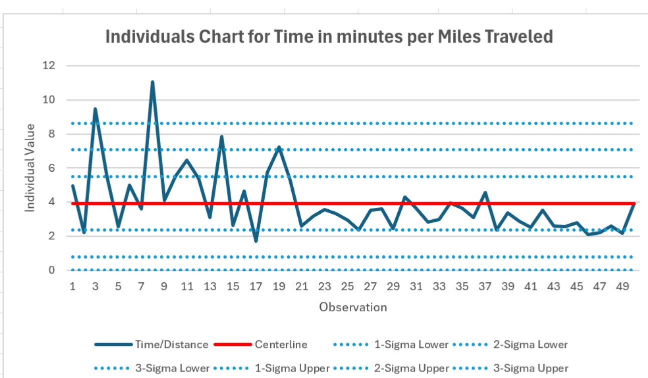
Analysis

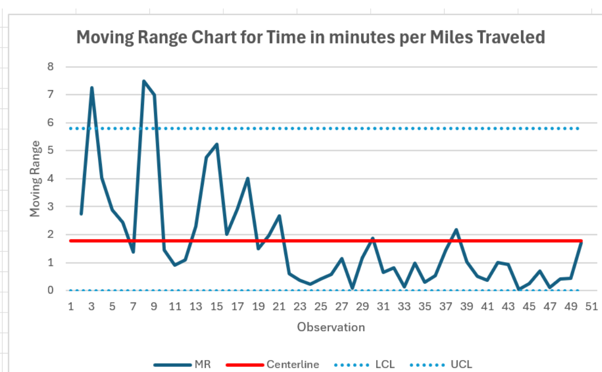
To analyze how strongly delivery times relate to distance, we made a linear regression model (shown below) including a line of best fit and its coefficient of determination (R^2). A coefficient of determination of 0.5926 suggests a moderately strong correlation between delivery distance and time. Given that most real-world regression models don't achieve high R^2 values such as 0.85 and above, a value of 0.5926 is fairly realistic, indicating that the simulated data does a decent job replicating real-world data.

To examine whether the time/delivery component is out of control, we looked at the two new columns in our dataset that were manually created. But first, we state the null and alternative hypotheses below for process stability:

- Null Hypothesis (H0): The Uber Eats delivery process is in control, with no significant deviations.
- Alternative Hypothesis (H1): The Uber Eats delivery process is out of control, with significant deviations.

Using the time/delivery and MR column, we can make the individual and moving range (MR) control charts (below) to assess process stability and identify variations.





The time/delivery variable is an excellent metric for delivery performance because it gives us a consistent way to measure delivery time for varying distances. When referencing the Western Handbook of Rules, the Individual chart violates the following:

1) A point plots outside the 3-sigma control limits (there are two)
2) Nine consecutive points are plotted on one side of the center line

We then reference the S/R Chart Guidelines for the MR chart, as the following violations are made:

1) 1 point at least 3 standard deviations above or below the centerline (there are three)
2) 9 points in a row, on the same side of the centerline

The violations seen in the individual and MR charts indicate an out-of-control process. This means that, based on the time/distance metric, we have inefficient delivery performance.

Informing Shareholders

We want to share our insights with the shareholders of Uber, which consists of the Uber Eats management, logistics teams, and delivery drivers themselves. We plan to communicate findings through reports and presentations.

Because of the various external factors drivers encounter, it’s difficult to eliminate inefficiencies in delivery times. However, several steps can be taken to improve the process. Enhancing navigation systems is a crucial starting point, as Uber has a history of map inaccuracies. Uber itself stated that “GPS data is noisy in urban environments and not sufficient to understand precise interactions between delivery partners and restaurants.” (Uber Technologies, 2018). This means investing in ways to improve algorithms for routes and potentially working with reputable navigation system companies. Another suggestion would be an in-app feature that allows drivers to leave location-based notes. This enables drivers to share important information about incorrect directions or optimal parking spots, helping future deliveries proceed more smoothly in that location. Lastly, Uber can implement new training modules that teach strategies to optimize delivery efficiency.

Conclusion

Uber Eats management, logistics teams, and delivery drivers will be impacted by this process because inefficient delivery times can lead to unhappy customers who may not choose to use Uber Eats in the future. Our analysis suggests that there are inefficiencies in delivery times, stressing the need for strategic change to counter the out-of-control process. To improve delivery performance, we recommend that Uber invest in more accurate navigation systems, introduce an in-app feature for drivers to post location-based notes, and implement training programs for efficient delivery strategies. An Out-of-Control Action Plan (OCAP) could be systematically designed to address delivery delays and achieve stability in the process. While our simulation provides useful insights, the following improvements could be made for more accurate insights:

- Generate a larger number of observations for more precise estimates
- Identify and assign more accurate probabilities and distributions to better reflect real-world Uber delivery scenarios
- Implement additional variables such as order size and food preparation time, which also impact delivery performance

References:

Boysen, N., Fedtke, S. & Schwerdfeger, S. Last-mile delivery concepts: a survey from an operational research perspective. OR Spectrum 43, 1-58 (2021). https://doi.org/10.1007/s00291-020-00607-8

Reyes, D., Savelsbergh, M., & Toriello, A. (2017). Vehicle routing with roaming delivery locations. H. Milton Stewart School of Industrial and Systems Engineering, Georgia Institute of Technology.

Uber Technologies Inc. (2018). How Trip Inferences and Machine Learning Optimize Delivery Times on Uber Eats. https://www.uber.com/blog/uber-eats-trip-optimization/