**Final Project**

**DIVE analysis – Customer and Market Analysis**

**Maria Arroyo**

**DISCOVER – Identify Customer Segments, Patterns and Dynamics**

1. **Identify customer segments, patterns, and dynamics.**

We use the data to identify the key characteristics of each segment. The goal is to find patterns that suggest a trip's purpose. Gemini helped us locate 10 different customer segments: Late-Night Traveler, Other, Evening Commuter, Errand Runner, Morning Commuter, Airport Traveler, Social Traveler, Casual Shopper, and Business Traveler. The results show a detailed nuanced view of the customer base.

* **Commuters are a High-Volume, Mid-Value Market**: The Evening Commuter and Morning Commuter segments are in the top 5 by trip volume. They represent a consistent, high-frequency customer base. The average fare is modest (around $10.50), but the sheer volume of trips makes this a critical part of the taxi business.
* **Late-Night Travelers are a Huge Opportunity**: The Late-Night Traveler segment is at the very top, with over 6.3 million trips. The average trip is slightly longer than a commuter trip and has a slightly higher fare, making it a very valuable target for acquisition and retention strategies.
* **Errand Runners are the Low-Value, High-Volume Base**: The Errand Runner segment is also a high-volume group (over 4 million trips), but with the lowest average fare (around $7.96) and shortest average trip duration. This is the "bread and butter" segment that relies on quick, convenient trips.
* **High-Value Segments are Less Frequent but Profitable**: The Airport Traveler, Business Traveler, and Tourist segments are lower in total trips, but their metrics are much higher.
  + Airport Travelers have the highest average fare ($42.27) and duration (39.21 minutes).
  + Business Travelers have a high average fare ($28.45) and are a great source of revenue.
  + Tourists have a high average fare ($27.67) and the largest average passenger count, confirming they travel in groups.
* **Social Travelers and Casual Shoppers are Mid-Value Segments**: These segments show moderate trip counts and average fares, indicating they are important but not as critical as the high-volume or high-value groups.

1. **Which pickup/drop-off zones show highest and lowest demand?**

* Highest demand pick-up zone: JFK
* Lowest demand pick-up zone: Staten Island Eltingville/Annadale/Prince's Bay
* Highest demand drop-off zone: Manhattan Upper East Side North
* Lowest demand drop-off zone: Staten Island Great Kills Park

1. **How do trip distances, durations, and purposes vary by time and location?**

* **Airport Traveler**: These trips continue to be a high-value segment, with high average fares and long trip durations. For example, a trip on a Sunday at midnight from location 132 (presumably JFK) averages a fare of $43.63 and a duration of 27.44 minutes. This confirms that airport trips are a significant source of revenue, justifying their own segment.
* **Late-Night Traveler**: This segment is the most dominant in the data provided, with a very high total trip count. Trips are generally short-to-medium distance and have a lower average fare compared to airport travelers. For example, at midnight on a Sunday, trips from location 100 average $12.50 with a distance of 2.88 miles. This suggests this segment is crucial for volume-based revenue, especially in high-density areas.
* **Social Traveler**: This segment, active on weekends (as seen from day 1 data, which corresponds to Sunday), is characterized by medium-distance trips with moderate fares. For instance, a trip from location 14 at 9:00 PM on a Sunday has an average fare of $18.05. This segment likely represents trips to and from bars, restaurants, and other entertainment venues.
* **Business Traveler**: Although not explicitly in the snippet, we can see the logic for this segment in the code. These are expected to be high-fare, weekday trips to key Manhattan locations, representing a premium, recurring customer base.
* **Errand Runner & Casual Shopper**: The data includes examples of these segments during midday on weekdays. They represent short, frequent trips for daily activities. For example, on a Thursday at 11 AM, trips from location 193 classified as "Errand Runner" have a short average distance of 0.83 miles and a quick duration of just over 6 minutes. These segments, while individually low-fare, contribute to a steady, predictable stream of revenue.

**INVESTIGATE – Drivers Behavior**

1. **Why are certain customer segments declining faster than others?**

* Commuters: The issue is speed and convenience, not necessarily the quality of the trip itself. They are willing to pay for a faster, more predictable service, which is where rideshare apps have an advantage. Rideshare apps excel at providing a sense of immediacy and efficiency. They can often provide a driver in a less-congested location, and the user-friendly app interface gives the perception of a faster, more streamlined experience from booking to payment.
* Airport Travelers: The issue is potentially overall satisfaction. The low tip percentage suggests a disconnect between the service provided and the customer's expectations, making them a vulnerable segment. The low tip percentage suggests a disconnect between the premium price of an airport trip and the customer's perceived value of the service. They may have experienced a less-than-ideal ride, felt the fare was too high, or simply prefer the transparent pricing and convenience of a rideshare app, making this valuable segment vulnerable to attrition.
* Business Traveler: Reflects a very low average tip percentage (15.71%). This is the lowest of all segments, which is a significant red flag. Business travelers are often on tight schedules and require seamless, reliable service. The low tip percentage suggests they are not satisfied with the current service. They may perceive it as less professional, less efficient, or simply not as integrated into their corporate workflow as a rideshare app, leading them to switch services.
* Rideshare Advantage: Rideshare apps excel at providing a sense of immediacy and efficiency. They can often provide a driver in a less-congested location, and the user-friendly app interface gives the perception of a faster, more streamlined experience from booking to payment.

1. **Is there unmet demand in areas/times not well served by the current fleet?**

* **High-Volume Segments**: The data shows that the Evening Commuter segment has the highest number of trips. The trips are concentrated on weekdays between 4 PM and 7 PM. The average fare and trip distance for this segment are consistently low, with an average fare of approximately $10-$11 and an average trip distance of around 2 miles. This suggests a high volume of short, reliable trips, which is a great base for the business.
* **Emerging Late-Night Demand:** The Late-Night Traveler segment is the second most common, with a large volume of trips from 9 PM to 4 AM. This segment has a higher average fare (around $11-$12) and longer average trip distances (around 2.5-3 miles) compared to commuters. This indicates that these trips are less about short, routine commutes and more about longer, less predictable journeys. This could represent a key area for growth.
* **Untapped Opportunities**: The "Other" segment, which represents unclassified trips, has a significant number of trips with a very wide range of trip distances (min and max) and variable fare amounts. This suggests a large, diverse set of trips that don't fit into the predefined commuter, tourist, or airport traveler categories. The maximum trip distance in this category is often very high (e.g., up to 99 miles), which suggests that many of these trips are likely long-distance or unusual in nature. This segment could represent a source of unmet demand. The high max\_trip\_distance values indicate that there may be a lot of longer trips that are currently unclassified, and that the company could be missing opportunities to serve these customers.
* **Commuter Demand:** The Morning Commuter and Errand Runner segments also show very high trip counts, indicating strong demand for short, functional trips during the daytime. These segments have low average fares and very short average trip distances (around 1-2 miles). The min\_trip\_distance and max\_trip\_distance for these segments are consistently low, reinforcing the idea that these are short, predictable trips.

1. **How do customer patterns differ from rideshare dominance zones?**

* **Late-Night Travelers:** This is the dominant segment in both high-demand and low-demand zones, but with a significant difference. In low-demand areas, late-night trips are much longer, with an average distance of 3.94 miles and an average fare of $16.52. In high-demand zones, these trips are shorter, with an average distance of 2.8 miles and a fare of $11.68. This indicates that late-night riders in low-demand areas are taking longer, more expensive trips, suggesting a reliance on taxis for journeys that might be less common or difficult to arrange via other means.
* **Commuters:** Commuter segments in low-demand zones have longer average trips and higher fares compared to their high-demand counterparts. For example, Morning Commuters in low-demand zones travel an average of 3.18 miles with a fare of $15.29, versus 2.01 miles and $10.3 in high-demand zones. This suggests that people in these zones are using taxis for less routine commuting, possibly due to a lack of frequent, affordable public transportation options.
* **Business Travelers & Tourists:** These high-value segments are noticeably smaller in low-demand zones. This suggests that these zones may not have as many business hubs or tourist attractions, or that rideshares have successfully captured these markets.

**VALIDATE – Test Segmentation and Behavior Analysis**

1. **Could observed trends be due to pandemic aftereffects or macroeconomic shifts?**

* The pandemic's shift to remote work likely caused a permanent change in commuting behavior, and the competition from rideshare services, a key macroeconomic shift, is likely capturing the trips that are still happening. The taxi fleet is not well-positioned to serve this new reality for commuters.
* In contrast, the data analysis showed that Airport Travelers and Tourists, while smaller in volume, are high-value customers who take longer, more lucrative trips. These segments were not as negatively impacted by the same factors that hit commuters. Airport travel, while recovering, is a captive market with a different set of priorities (e.g., luggage space, fixed rates), and tourist travel is less about daily commute and more about long-distance, group trips. This is where the service still performs well and can remain competitive.
* The analysis shows that while the total number of trips and total revenue have dropped significantly from 2019 to 2022, the average values per trip have increased. This suggests a fundamental shift in the business.
* The decrease in total trips and total amount is likely a direct result of the pandemic's impact on travel and commuting habits. With more people working from home and a general reduction in movement, the overall volume of taxi trips has plummeted.
* In summary, the business has moved from a high-volume, lower-value model to a lower-volume, higher-value model. The total pie is smaller, but each slice is bigger. This confirms that to grow revenue, the company should not only try to recover lost volume but also focus on attracting and retaining these higher-value, longer-distance customers.

1. **Are “declines” just shifts to other taxi service types not in the dataset**

* Evidence from Low-Demand Zones: The analysis of High-Demand vs. Low-Demand zones revealed that in low-demand areas, taxi trips are less frequent but are on average longer and more expensive. This suggests that for quick, short trips in these areas, customers are opting for other services, likely rideshare apps.
* The Case of JFK Airport: We also found that while JFK Airport (Location ID 132) is a massive source of taxi pickups, it is not a top drop-off location. This indicates a key "one-way" pattern of travel, likely with people arriving at the airport via a rideshare service and then taking a taxi. This one-way trend is a strong indicator of a shift to other services, with the fleet still holding a competitive advantage in a specific type of trip from the airport.

**EXTEND – Acquisition/Retention Strategies**

1. **What targeted marketing or loyalty programs could win back high-value customers?**

* Offer a flat-rate, pre-booked fare for common commuter routes during peak hours. This removes the uncertainty of cost, a major pain point for commuters.
* Since commuters have a high card-payment percentage, lean into technology. Ensure a seamless app experience with quick booking, in-app payments, and clear route visualization.
* For the business travelers, the taxi company could focus on corporate accounts and seamless billing. Offer a "Business Plus" loyalty program where every five trips earns the rider an upgrade to a premium vehicle, or a free trip to or from a key business district.

1. **Which geographic areas should be prioritized for re-entry or expansion?**

* The data showed that airport travelers, while these are high-value customers, their tip percentage is the lowest, suggesting potential dissatisfaction. The taxi company should consider creating a loyalty program for frequent airport travelers
* The data suggests that tourists are a high value segment. We could keep and increase tourist retention by increasing the visibility of the service and simplifying accessibility. The taxi company could offer bundle packages for tour groups and create partnerships with hotels and tourist attractions.