In this project, I will analyze the NYC Taxi Yellow data to draw meaningful insights and visualize through Power BI

NYC Taxi Project

Minh Dang

Contents

A) Introduction:

In this project, I will demonstrate my understanding about Data Analyst using Microsoft Fabric to transform data. With the processed data, I will load it into Power Bi to visualize and show my observations. These are the goals that I want to achieve in this project:

- The popularity and distribution of each vendor.
- Special patterns within payment methods.
- Total distance travelled and total money paid.
- The customer behaviours and well-known zones.

Using those valuable insights, users can draw significant conclusions and decisions to rearrange the taxi drivers and cost to increase the benefits and reduce the risks.

B) Datasets and Tools:

1) About the datasets:

The datasets contains yellow trip records in 2024, which include pickup and drop-off dates/times, pickup and drop-off locations, trip distances, itemized fares, rate types, payment types and driver-reported passenger counts. The dataset was collected and provided to NYC Taxi and Limousine Commission (TLC) by technology providers authorized under the Taxicab & Livery Passenger Enhancement Programs (TPEP/LPEP). The data contains parquet files of yellow trip recorded within each month followed by a taxi zone lookup table. Furthermore, there was also a taxi zone lookup file that helps explaining the location id with borough, zone and service zone.

The detail for each field of the parquet file is shown as below.

Field Name	Description
VendorID	A code indicating the TPEP provider that provided the record. 1 = Creative Mobile Technologies, LLC 2 = Curb Mobility, LLC 6 = Myle Technologies Inc 7 = Helix
tpep_pickup_datetime	The date and time when the meter was engaged.
tpep_dropoff_datetime	The date and time when the meter was disengaged.
passenger_count	The number of passengers in the vehicle.
trip_distance	The elapsed trip distance in miles reported by the taximeter.
RatecodeID	The final rate code in effect at the end of the trip. 1 = Standard rate 2 = JFK 3 = Newark 4 = Nassau or Westchester 5 = Negotiated fare 6 = Group ride 99 = Null/unknown
store_and_fwd_flag	This flag indicates whether the trip record was held in vehicle memory before sending to the vendor, aka “store and forward,” because the vehicle did not have a connection to the server. Y = store and forward trip N = not a store and forward trip
PULocationID	TLC Taxi Zone in which the taximeter was engaged.
DOLocationID	TLC Taxi Zone in which the taximeter was disengaged.
payment_type	A numeric code signifying how the passenger paid for the trip. 0 = Flex Fare trip 1 = Credit card 2 = Cash 3 = No charge 4 = Dispute 5 = Unknown 6 = Voided trip
fare_amount	The time-and-distance fare calculated by the meter. For additional information on the following columns, see https://www.nyc.gov/site/tlc/passengers/taxi-fare.page
extra	Miscellaneous extras and surcharges.
mta_tax	Tax that is automatically triggered based on the metered rate in use.
tip_amount	Tip amount – This field is automatically populated for credit card tips. Cash tips are not included.
tolls_amount	Total amount of all tolls paid in trip.
improvement_surcharge	Improvement surcharge assessed trips at the flag drop. The improvement surcharge began being levied in 2015.
total_amount	The total amount charged to passengers. Does not include cash tips.
congestion_surcharge	Total amount collected in trip for NYS congestion surcharge.
airport_fee	For pick up only at LaGuardia and John F. Kennedy Airports.
cbd_congestion_fee	Per-trip charge for MTA's Congestion Relief Zone starting Jan. 5, 2025.

2) Microsoft Fabric:

Microsoft Fabric is an end-to-end analytics platform that unifies data engineering, data science, real-time analytics and business intelligence under one roof. It integrates services like Power BI, Data Factory and Synapse into a single SaaS experience, enabling seamless data movement, transformation and visualization. Designed for scalability and collaboration, Fabric empowers organizations to build robust data pipelines, model complex datasets and deliver impactful insights—all without switching between fragmented tools.

When it comes to processing Parquet files, Microsoft Fabric supports them across various components, especially within Data Factory pipelines and Dataflow Gen2. You can configure

Parquet as both a source and destination format in copy activities, with options to set compression types (e.g., gzip, snappy, Brotli), control file naming patterns and optimize write performance using V-Order. In Dataflows, Fabric uses Power Query connectors to ingest Parquet files, allowing for transformation and enrichment before loading into Lakehouses or other destinations. This makes it ideal for handling large-scale, columnar data efficiently across your analytics workflows.

C) Processing in Microsoft Fabric:

1) Dataset Stage Processing Pipeline:

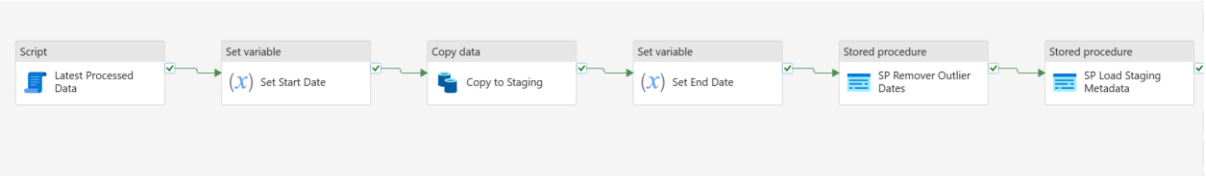


Figure 1: NYC Stage Processing Pipeline.

The picture above demonstrates the steps to load data from the Lakehouse. The process include five steps from getting the latest processed date to load staging metadata.

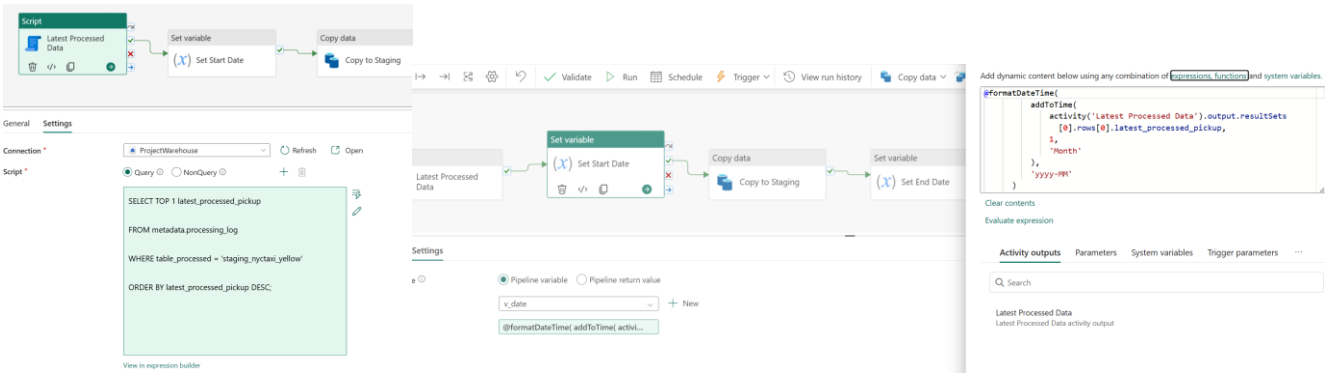
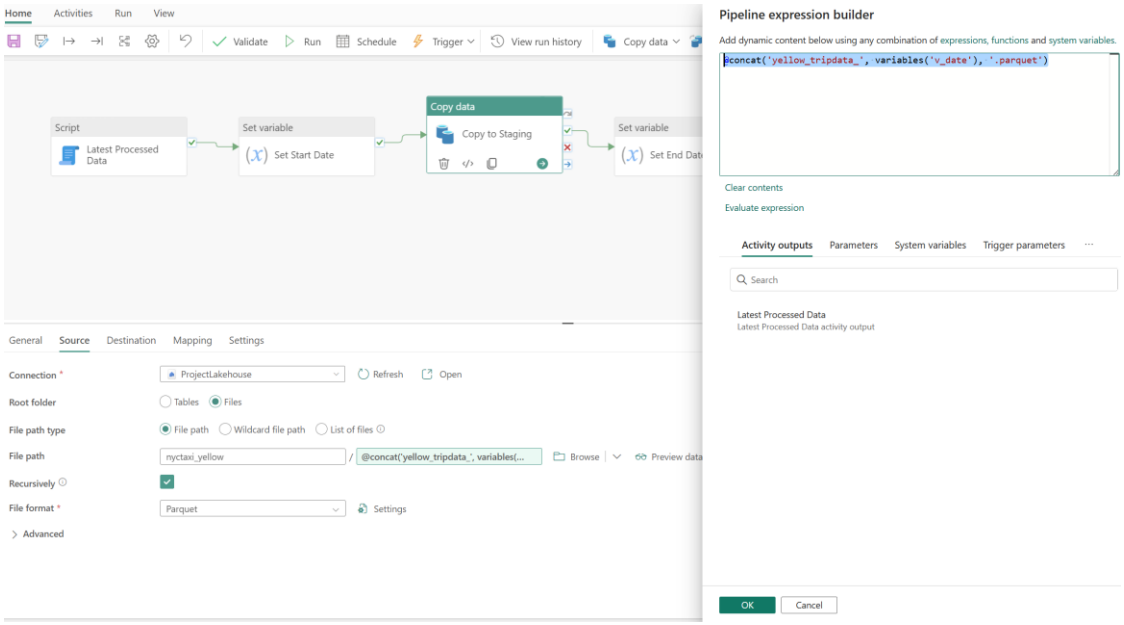


Figure 2: Obtain the latest processing date and Set Start Date



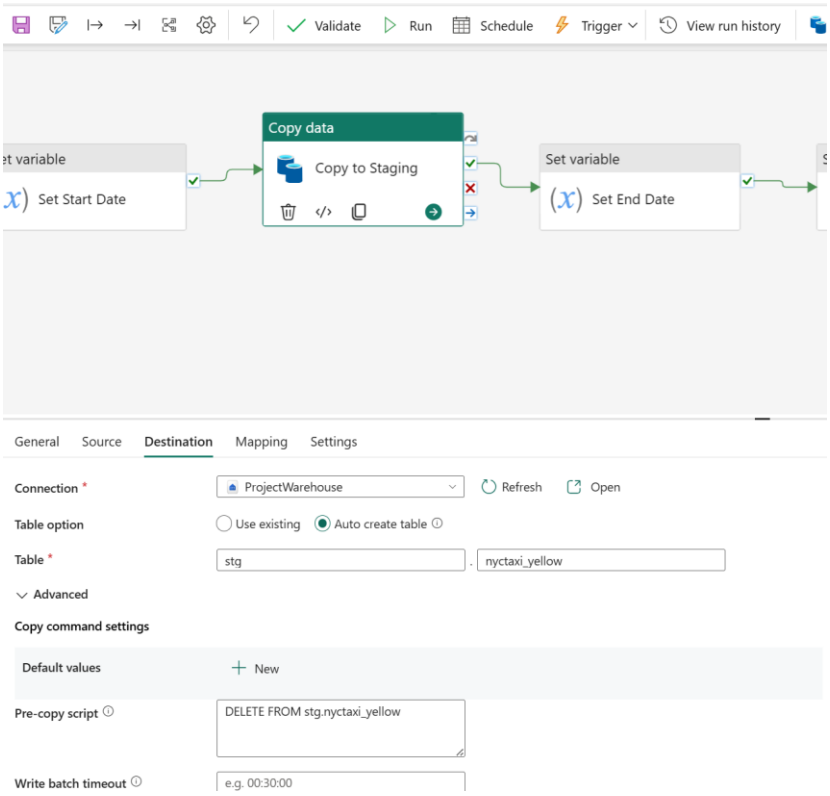


Figure 3: Copy Data to Staging.

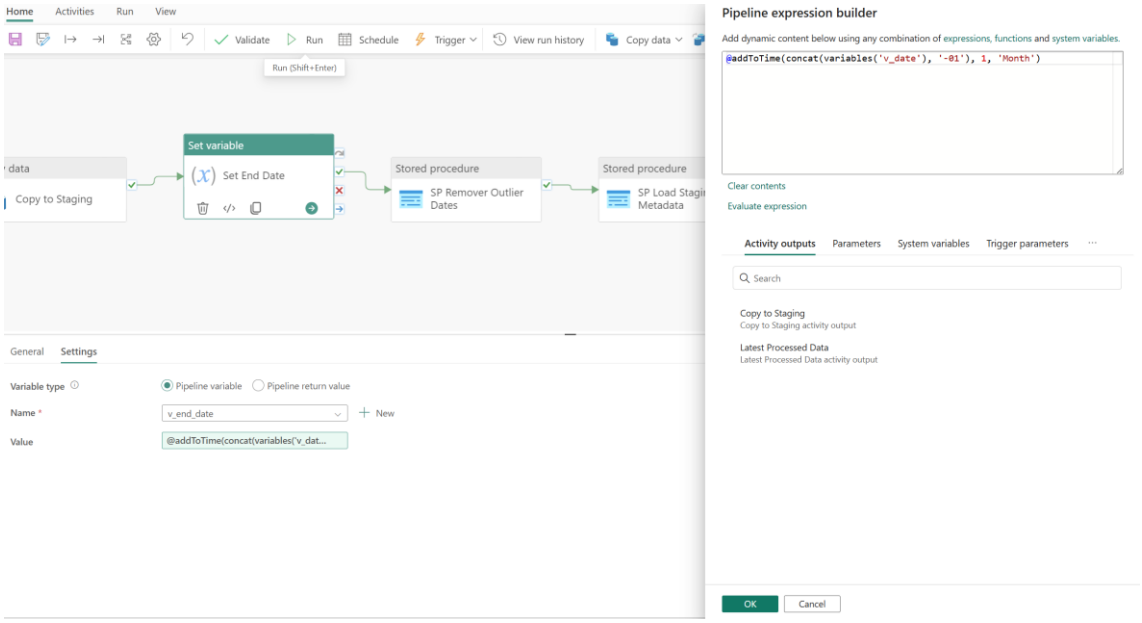


Figure 4: Set End Date.

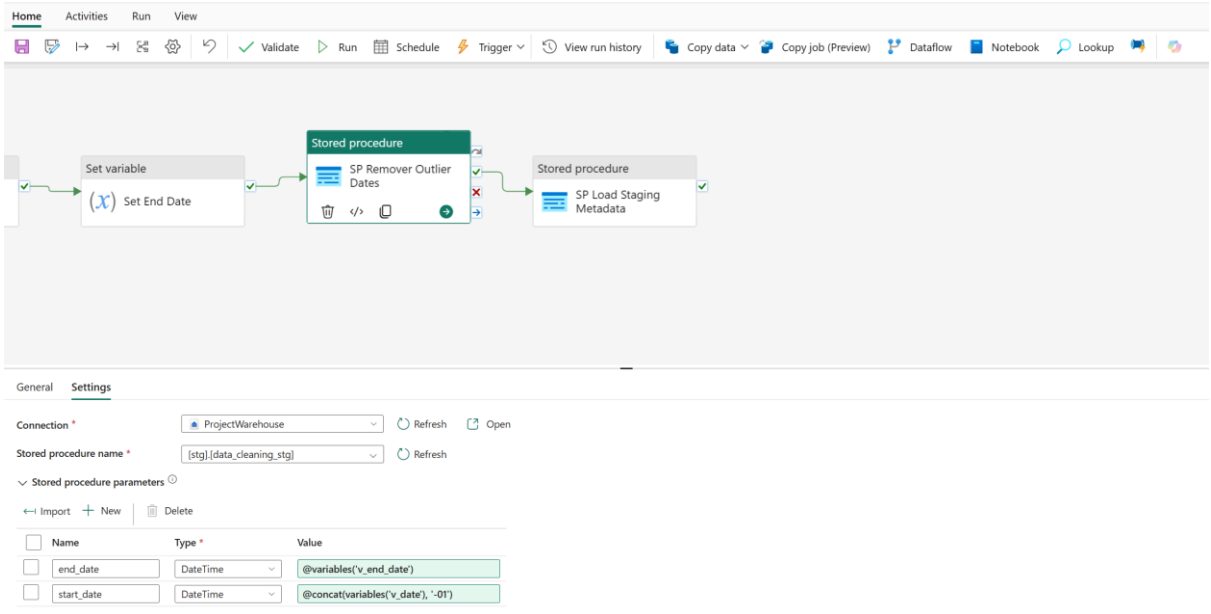


Figure 5: Remove outlier Date.

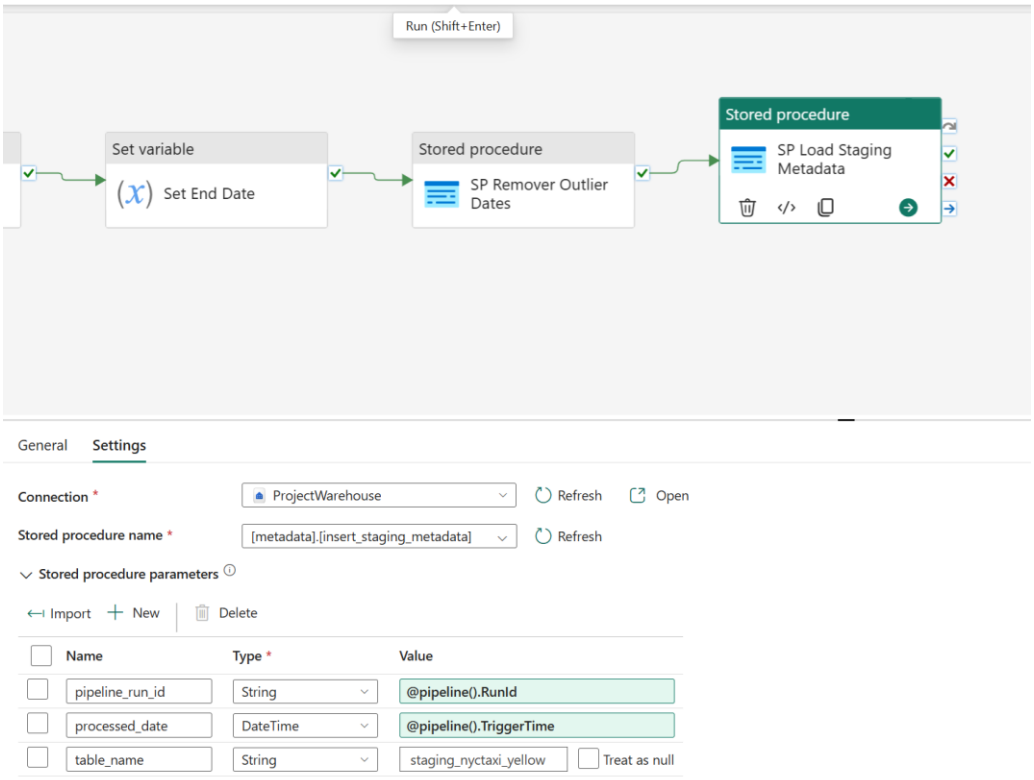


Figure 6: Load Staging Metadata.

In the first two steps, the latest processing entry based on timestamp for table staging nyctaxi was extracted, followed by adding one month to that date and converting into full string. Once the start date was defined and stored as v_date, the next step will start copying the files from nyctaxi_yellow using the prefix “yellow_tripdata_” combining with increased month and end with “.parquet” to a table called “stg.nyctaxi_yellow”. Then, the end date will be set at the start of the next month to help the next procedure removing the outlier date between start and end date. Finally, the pipeline run id, processing data and table name were recorded for later monitors.

2) Copy Lookup Table:

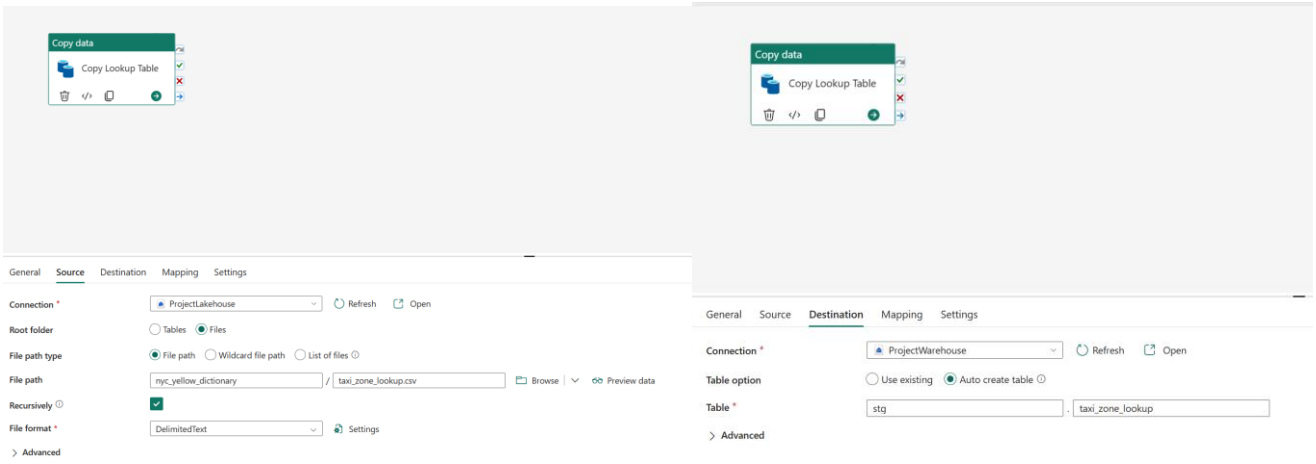


Figure 7: Copy Lookup Table

As mentioned above, the taxi zone lookup helped me converting the location id into appropriate borough, zone and service zone. Therefore, copying the file from lake house to warehouse for processing is mandatory.

3) Presentation process:

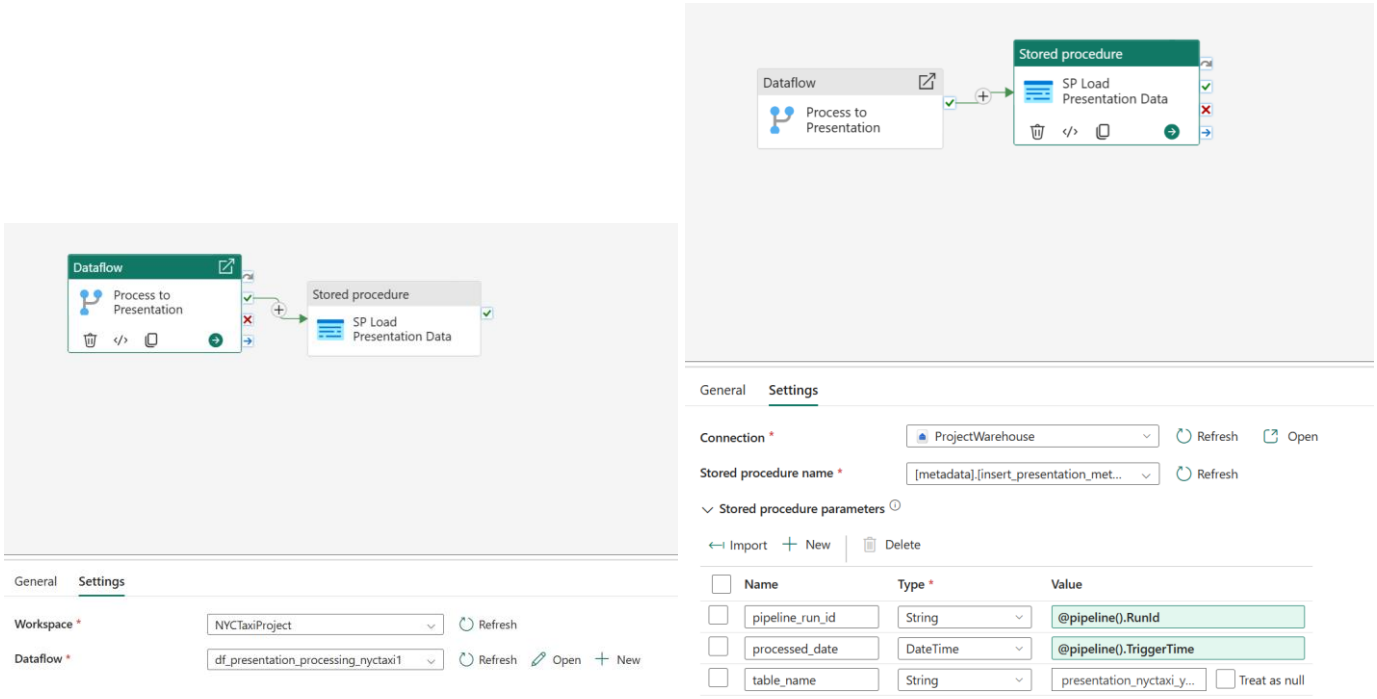


Figure 8: Process Data for presentation.

Before loading the nyc taxi data for presentation, another log was recorded include run id, process data and table name.

4) SQL Processing code:

The SQL code will be attached along with this report. In summary, these are the steps that I have made:

- Retrieve the earliest and latest pickup date from the staging dataset for checking.
- Defined a store procedure to remove outlier records outside valid data range using start and end date and then check again as above.
- Create a metadata schema to store table processing log that records the pipeline run id, table processed, row processed, latest process date time in data and process datetime.
- Generate a manual snapshot for testing.
- Insert metadata staging for logging.
- Drop and recreated if have a nyctaxi_yellow table.
- Transform staging data into presentation format, including mapping vendor id to names, format dates, joins with zone lookup table and translate payment type.
- Insert presentation metadata.
- View all processing log.

5) Final Transformation steps in Power Query:

After merging lookup table with the main nyctaxi dataset, I removed unwanted columns and reformat some dataset, followed by converting the Payment type and Vendor into its original meaning. Below was my final dataset that was ready for presentation:

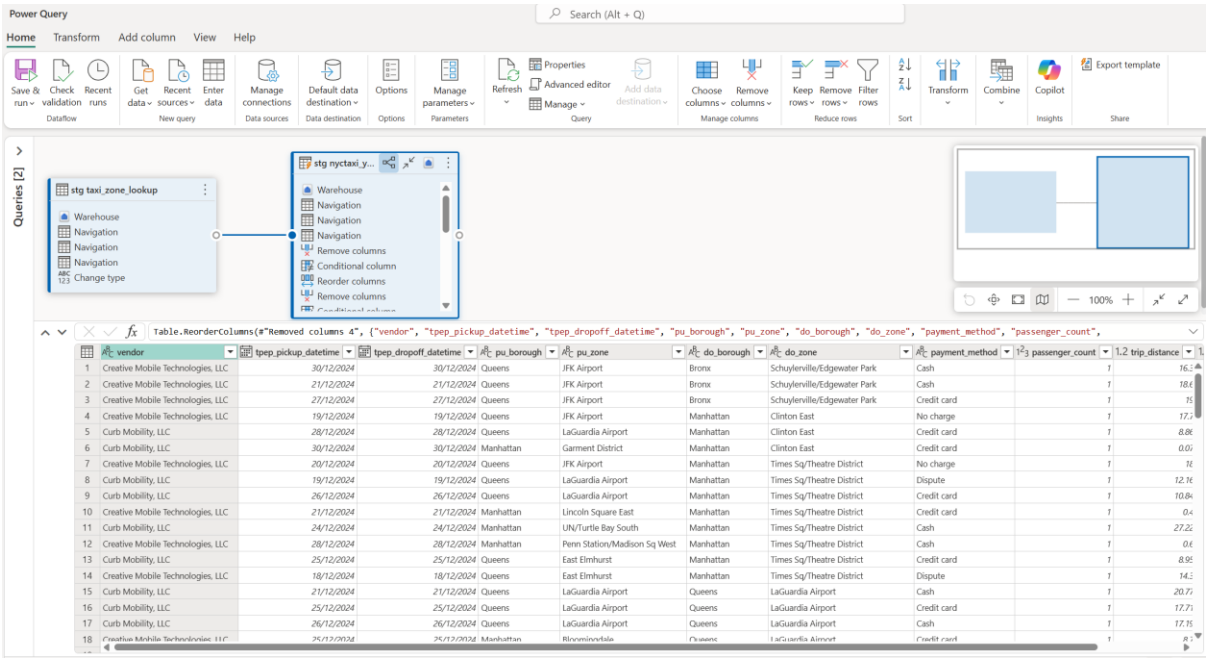


Figure 9: Final Dataset.

D) Power BI Import and Visualization:

1) Import Data:

Using the SQL Endpoint, I connected to project Warehouse in Microsoft Fabric. The link can be taken in here and paste it under “SQL Server” section in Power BI.

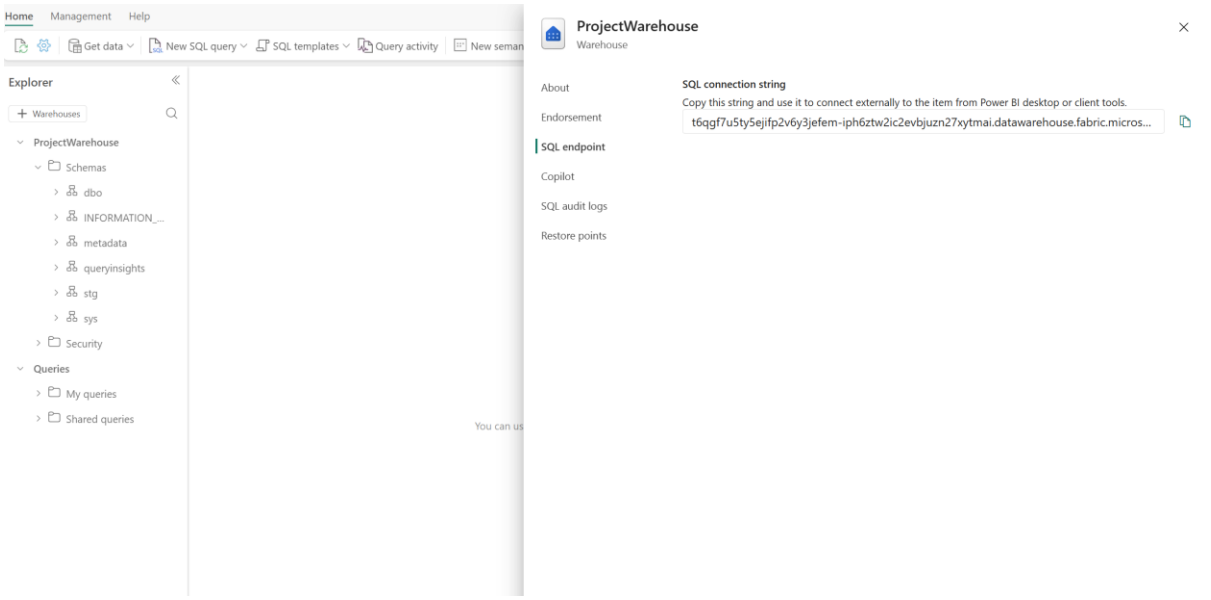


Figure 10: SQL Endpoint Location.

2) Performance Overview:

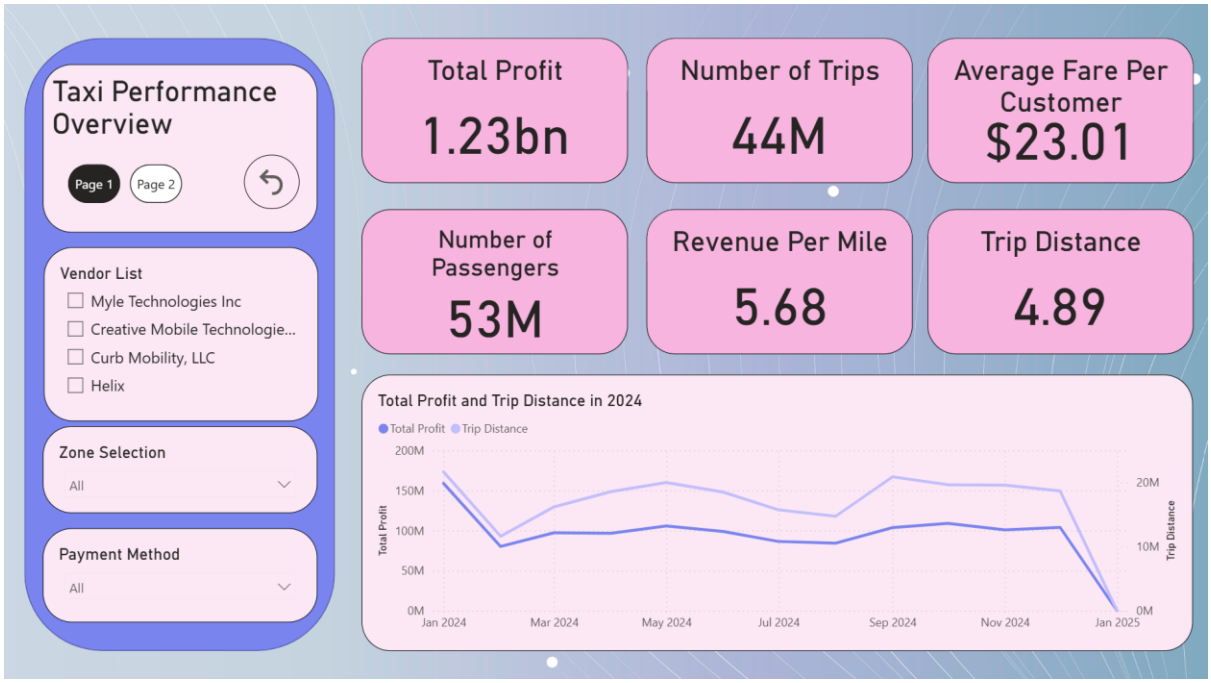


Figure 11: Overview Dashboard.

The 2024 NYC taxi performance dashboard reveals a robust operational landscape, with 44 million trips generating a total profit of $1.23 billion and transporting 53 million passengers. The average fare per customer stands at $23.01, while the average trip distance is 4.89 miles, yielding a revenue per mile of $5.68—indicative of healthy fare efficiency across zones. Temporal trends in the line chart suggest a seasonal dip in trip distance mid-year, while profit remains relatively stable, implying potential optimization in route planning or pricing strategies. To enhance profitability and customer satisfaction, taxi businesses should consider dynamic pricing models tied to trip distance and demand patterns, incentivize digital payments for operational efficiency and explore zone-based service enhancements. Additionally,

integrating predictive analytics to forecast peak demand periods and adjusting fleet deployment accordingly could further streamline operations and reduce idle time.

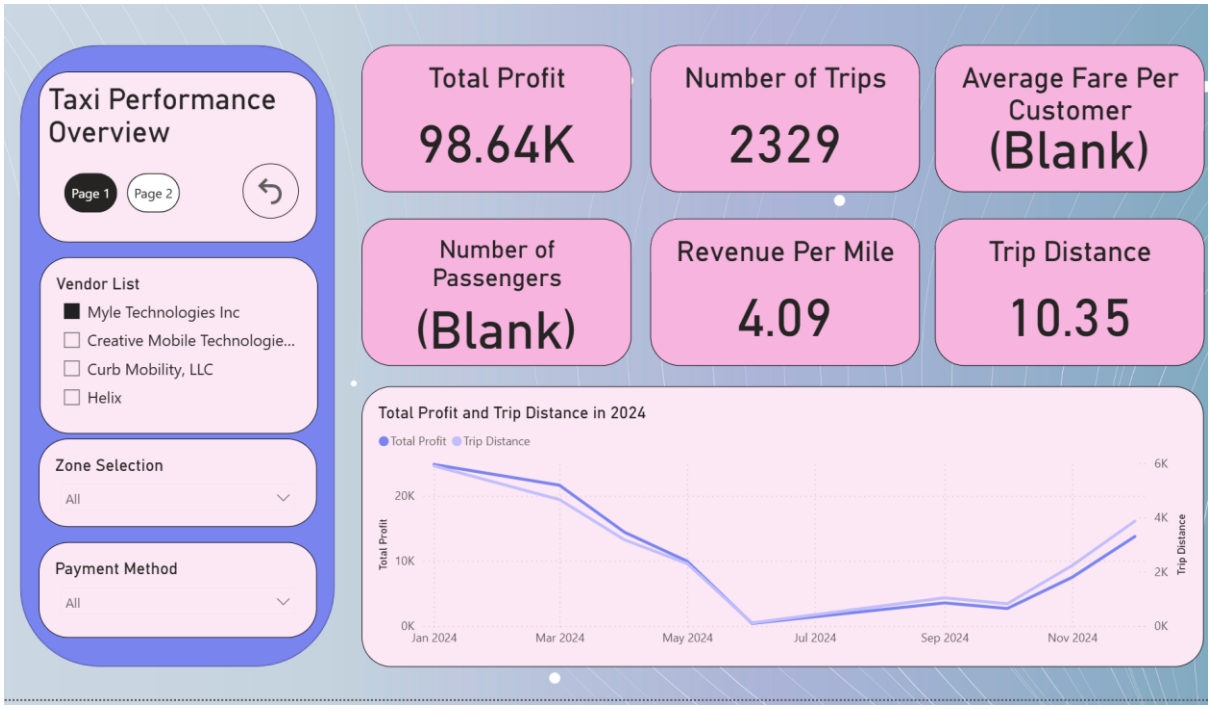


Figure 12: Myle Technology.

For Myle Technologies Inc. from the 2024 NYC Taxi Performance dashboard, the result reveals a nuanced operational profile that warrants both strategic reflection and targeted improvement. With **2,329 trips** generating **$98.64K in total profit**, the vendor shows a relatively modest footprint compared to larger players. However, the **average trip distance of 10.35 miles**—significantly above city norms—suggests that Myle may be servicing longer-haul routes, potentially in outer boroughs or airport corridors. The **revenue per mile of $4.09**, while slightly below the benchmark seen in broader datasets (e.g., $5.68 from other vendors), still indicates a viable pricing structure, though possibly under-optimized for distance.

The absence of **passenger count data** is a critical gap. Without this metric, it's impossible to assess load efficiency, average occupancy or fare-per-passenger ratios—key indicators for operational health and profitability. This omission could stem from data integration issues, inconsistent logging practices or system-level limitations in Myle's dispatch or payment infrastructure. Rectifying this should be a top priority, as it impairs both internal performance tracking and external benchmarking.

Strategic Recommendations:

- **Data Completeness Audit**: Myle should immediately investigate the missing passenger data. Implementing stricter ETL validation and metadata logging can ensure completeness and consistency across reporting layers.
- **Zone-Level Profitability Analysis**: Given the longer trip distances, Myle may benefit from a zone-based profitability breakdown to identify high-margin corridors and optimize fleet allocation.

- **Fare Optimization**: With a lower revenue per mile, Myle could explore dynamic pricing models or bundled fare strategies for longer trips to improve margins.
- **Vendor Benchmarking**: Comparing performance against peers like Curb Mobility or Awesome Mobile Technologies—especially in similar zones—can uncover pricing inefficiencies or service gaps.
- **Operational Transparency**: Enhancing dashboard granularity with metrics like average fare per customer and passenger count will improve stakeholder confidence and support strategic planning.

In short, Myle Technologies shows potential in niche service areas but must address data integrity and pricing strategy to unlock full performance visibility and profitability.

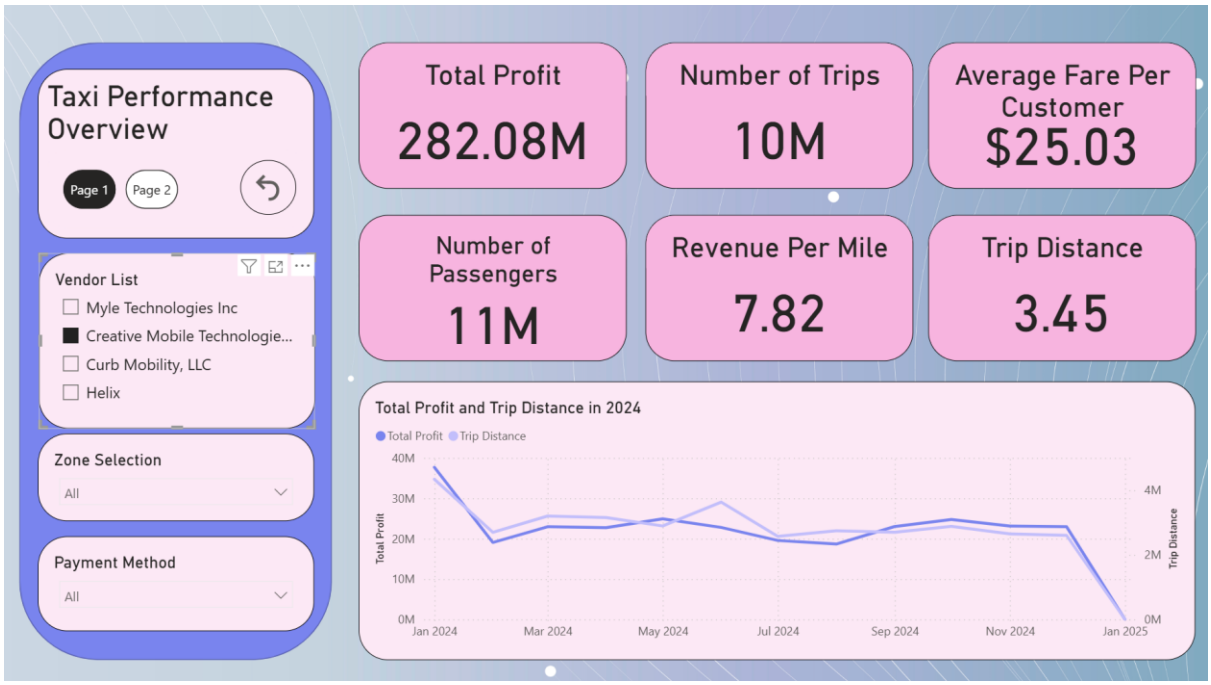


Figure 13: Creative Mobile Technologies.

Moving on to **Creative Mobile Technologies, Inc.,** from the 2024, the dashboard reveals a strong operational profile with several standout metrics. The vendor recorded **10 million trips**, generating **$282.08 million in total profit**, transporting **11 million passengers**, and achieving an **average fare per customer of $25.03**. These figures position Creative Mobile Technologies as a high-volume, high-efficiency operator within the city's taxi ecosystem. When benchmarked against the overall dashboard averages (e.g., 44M trips and $1.23B profit across all vendors), Creative Mobile Technologies accounts for roughly **23% of total trips** and **23% of total profit**, indicating proportional strength in both scale and profitability. However, what sets this vendor apart is its **revenue per mile of $7.82**, which is **significantly higher** than the broader average of $5.68. This suggests superior fare efficiency, possibly due to optimized routing, premium service zones, or effective fare structuring.

Interestingly, the **average trip distance is 3.45 miles**, which is **shorter than the overall average of 4.89 miles**, yet the vendor still achieves higher revenue per mile. This points to a strategic focus on **dense, high-demand urban zones** where shorter trips yield higher margins—such as Midtown Manhattan or airport feeder routes.

Professional Recommendations:

- **Zone-Level Optimization**: Given the high revenue per mile and shorter trip distances, Creative Mobile Technologies should continue to prioritize high-density zones. A heatmap analysis of trip origin-destination pairs could further refine fleet deployment.
- **Fare Structure Review**: The elevated fare per customer ($25.03) suggests room to explore tiered pricing models or bundled services (e.g., luggage handling, priority pickup) to enhance customer value perception.
- **Passenger Load Efficiency**: With 11M passengers over 10M trips, the average occupancy is approximately **1.1 passengers per trip**. This is slightly below optimal for shared ride models. Introducing or promoting multi-passenger ride options could improve per-trip profitability.
- **Temporal Trend Analysis**: The line graph shows consistent profit and trip distance trends throughout the year. However, identifying seasonal peaks (e.g., holidays, tourism surges) and aligning marketing or fleet expansion strategies accordingly could unlock additional revenue.
- **Benchmarking Against Peers**: Compared to vendors like Myle Technologies (lower trip volume, longer distances, and missing passenger data), Creative Mobile Technologies demonstrates superior data completeness and operational efficiency. Maintaining this edge through robust ETL pipelines and real-time analytics will be key.

In summary, Creative Mobile Technologies is a high-performing vendor with strong fare efficiency and strategic zone targeting. Continued investment in data-driven fleet management, customer segmentation, and service differentiation will help sustain and grow its competitive advantage.

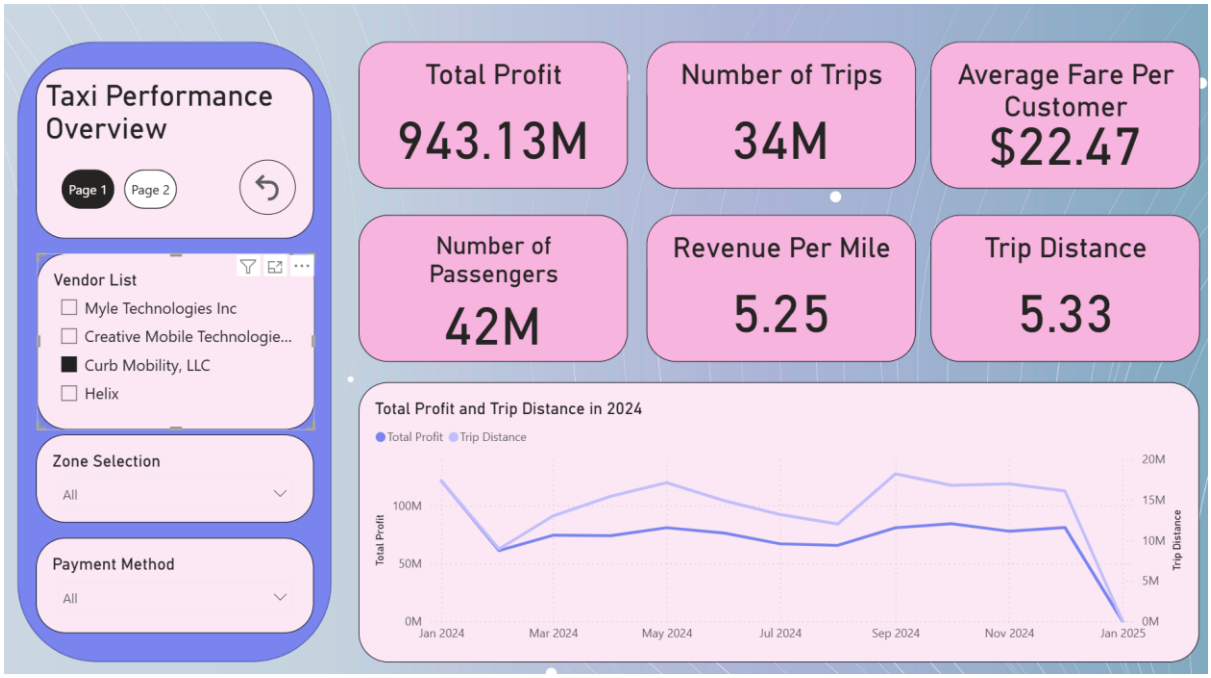


Figure 14: Curb Mobility.

Curb Mobility, LLC demonstrates a commanding presence in the NYC taxi landscape for 2024, with performance metrics that reflect both scale and operational maturity. The company logged **34 million trips**, transporting **42 million passengers**, and generated **$943.13 million in total profit**. These figures represent approximately **77% of the total trips** and **76% of the total profit** across all vendors analyzed, positioning Curb as the dominant player in both volume and revenue. The **average fare per customer** is **$22.47**, slightly below Creative Mobile Technologies' $25.03, yet Curb compensates with broader reach and higher passenger throughput.

From an efficiency standpoint, Curb's **revenue per mile of $5.25** and **average trip distance of 5.33 miles** are closely aligned with the overall market averages (e.g., $5.68 and 4.89 miles respectively), indicating a balanced pricing model that scales well across diverse zones. The **passenger-to-trip ratio of 1.24** suggests a healthy occupancy rate, potentially benefiting from shared ride models or high-frequency urban corridors.

Temporal Trend Observations: The line graph tracking monthly profit and trip distance shows consistent performance throughout the year, with no sharp volatility. This stability implies effective fleet management and demand forecasting, likely supported by robust dispatch algorithms and zone-based optimization.

Strategic Insights & Recommendations:

- **Fare Elasticity Modeling**: With slightly lower average fares, Curb could explore targeted fare adjustments in premium zones or during peak hours to boost per-trip profitability without sacrificing volume.
- **Zone Segmentation Strategy**: Given the scale, a deeper dive into zone-level profitability and trip density could uncover underperforming areas or high-margin opportunities for strategic fleet reallocation.
- **Customer Segmentation & Loyalty**: With 42M passengers served, Curb is well-positioned to implement loyalty programs or personalized promotions based on rider frequency, payment method, or trip patterns.
- **Operational Benchmarking**: Compared to Creative Mobile Technologies (higher fare per customer, shorter trip distance, and higher revenue per mile), Curb's strength lies in volume and consistency. A hybrid strategy that blends Creative's fare efficiency with Curb's scale could yield optimal results.
- **ETL & Data Governance**: The completeness and granularity of Curb's metrics suggest mature data pipelines. Continued investment in real-time analytics and predictive modeling will be essential for sustaining competitive advantage.

In summary, Curb Mobility's performance reflects a well-oiled operation with strong market coverage and reliable profitability. By fine-tuning fare strategies and leveraging its vast passenger base, the company can further solidify its leadership while unlocking new growth avenues.

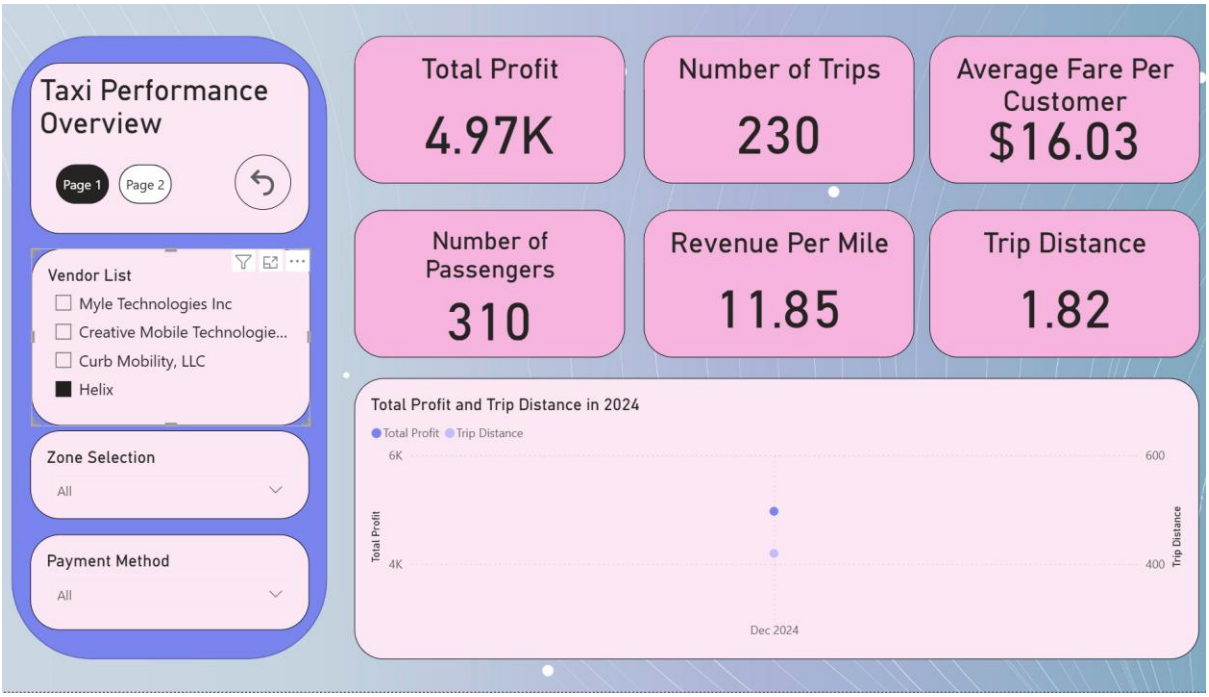


Figure 15: Helix Overview.

Helix appears to be a recent entrant in the NYC taxi market, with operational data limited to December 2024—a detail that strongly suggests either a late-year launch or a pilot phase. Despite the limited timeframe, the company has logged 230 trips, transporting 310 passengers, and generated $4.97K in total profit. While these figures are modest in absolute terms, they offer valuable signals about Helix's early positioning and strategic potential.

Key Performance Observations:

- **Average Fare Per Customer**: At **$16.03**, Helix's fare is notably lower than the market average (e.g., $22.47 for Curb Mobility), which may reflect a competitive pricing strategy aimed at rapid market penetration or service in lower-demand zones.
- **Trip Distance**: The average trip length is **1.82 miles**, suggesting Helix is operating in **dense urban areas** with short-haul routes—possibly Midtown, Downtown, or transit-adjacent zones.
- **Revenue Per Mile**: A standout metric at **$11.85**, which is **more than double** the market average (~$5.25–$7.82). This indicates **high fare efficiency**, likely driven by short trips with base fare premiums or surcharges (e.g., congestion pricing, airport fees).

Strategic Insights:

- **High Revenue Efficiency**: Despite low trip volume, Helix's revenue per mile suggests strong monetization of short-distance rides. This could be a deliberate strategy to target high-frequency, high-margin corridors.
- **Passenger Load**: With **310 passengers over 230 trips**, the average occupancy is **1.35 passengers per trip**, which is relatively high. This may reflect early adoption of shared ride models or group bookings.

- **Scatter Plot Positioning**: The December-only data point on the scatter plot places Helix in the lower-left quadrant—low profit, short distance—but with a steep revenue gradient. This reinforces the idea of **premium pricing on short routes**.

Recommendations for Growth:

- **Expand Temporal Coverage**: To validate performance trends, Helix should prioritize full-year data capture and ensure robust ETL pipelines for continuous reporting.
- **Zone-Level Expansion**: If current success is tied to specific zones, replicating the model in similar high-density areas could scale profitability.
- **Fare Structure Review**: While the revenue per mile is strong, the relatively low fare per customer may limit upside. Introducing tiered pricing or value-added services (e.g., priority pickup, in-ride amenities) could enhance margins.
- **Marketing & Brand Positioning**: As a new player, Helix should leverage its early efficiency metrics to position itself as a premium short-trip provider, possibly targeting business commuters or tourists.

In summary, Helix's December snapshot reveals a lean but promising operation with high fare efficiency and strategic zone targeting. With careful scaling and data-driven refinement, the company could carve out a profitable niche in NYC's competitive taxi ecosystem.

3) Region and Payment:

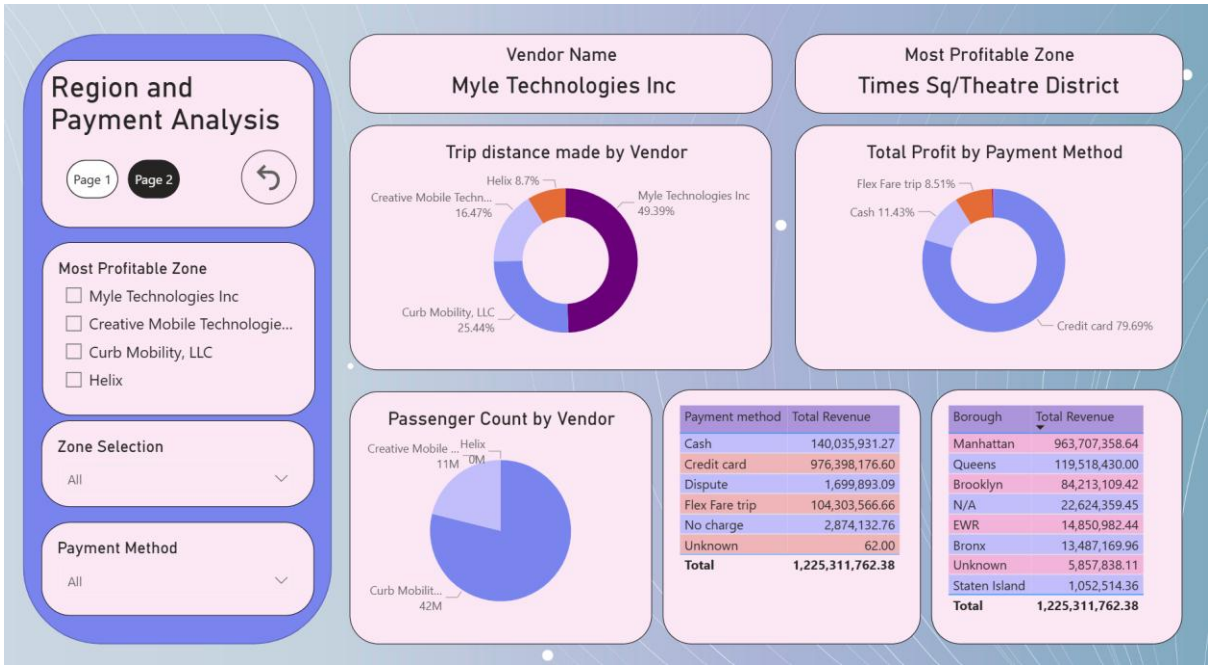


Figure 16: Region and Payment Overview

The dashboard offers a multidimensional view of NYC taxi operations, with a particular focus on vendor performance, geographic profitability, and payment behavior. One of the most striking insights is the dominance of **Curb Mobility, LLC**, which accounts for **over two-thirds of total profit (67.37%)**, **trip distance (68.26%)**, and **passenger count (67.37%)**. This

level of market share suggests not only operational scale but also strategic coverage across high-demand zones. In contrast, **Myle Technologies Inc.** and **Creative Mobile Technologies** trail behind with roughly 15–17% each across these metrics, indicating more niche or regionally concentrated service models.

From a regional perspective, **Times Sq/Theatre District** emerges as the most profitable zone, generating **$12.23 million** in revenue—outpacing other high-traffic areas like **Upper East Side South** and **Midtown Center**. This reinforces the economic importance of tourism and entertainment hubs, where short but frequent trips likely contribute to high cumulative revenue. Notably, **airport zones** such as **JFK** and **LaGuardia** also rank among the top earners, suggesting that long-distance, premium-fare trips continue to be a vital revenue stream. Vendors aiming to optimize profitability should consider dynamic fleet allocation strategies that balance short-haul urban density with long-haul airport connectivity.

The payment method analysis reveals a clear preference for **credit card transactions**, which account for **over $1.04 billion**—nearly **85% of total revenue**. This reflects both rider convenience and the city's push toward cashless mobility. **Cash payments**, while still significant at **$140 million**, are declining in relative share, and alternative methods such as **dispute**, **no charge**, and **voided trips** each contribute marginally (around $1.4 million). The uniform revenue values across these minor categories suggest either placeholder data or aggregated error handling, which may warrant a deeper audit of ETL processes to ensure accurate categorization and reporting.

Professionally, this dashboard underscores the importance of **zone-level profitability mapping**, **vendor benchmarking**, and **payment behavior segmentation**. Vendors like Myle Technologies could benefit from expanding into high-yield zones such as Times Square or Midtown, while also refining their digital payment infrastructure to align with consumer preferences. Moreover, the consistency in vendor share across profit, distance, and passenger count implies that operational efficiency—not just volume—is a key differentiator. Future enhancements to this dashboard could include time-of-day analysis, fare elasticity modeling, and predictive demand forecasting to support more granular decision-making.

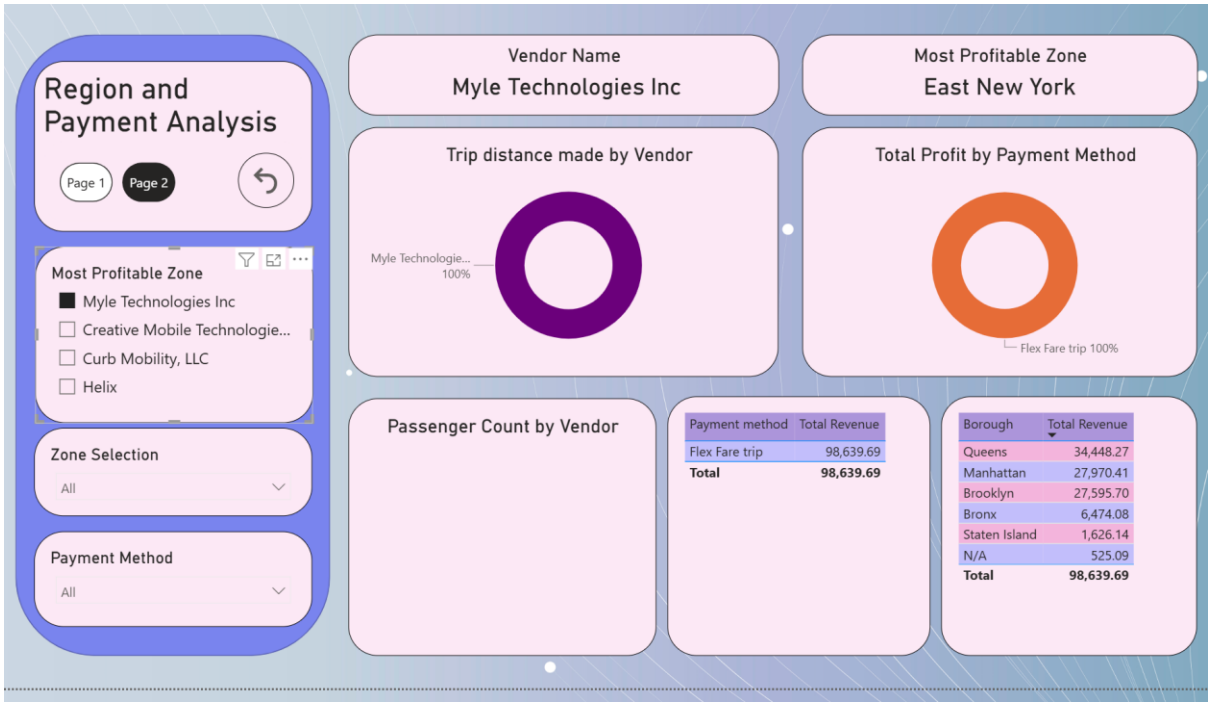


Figure 17: Myle Technologies detail Profit.

This dashboard demonstrated a focused snapshot of **Myle Technologies Inc.** within the NYC taxi ecosystem, with a distinct emphasis on regional profitability and payment behavior. The most profitable zone for Myle is **East New York**, a notable deviation from high-revenue zones like Times Square or Midtown seen in broader analyses. This implies that Myle may be strategically targeting underserved or outer-borough areas, potentially capturing demand where competition is lower and trip density is more predictable.

The payment method analysis is particularly revealing: **100% of total profit is attributed to "Flex fare trip"**, a non-standard category that likely reflects either promotional pricing, bundled services, or a custom fare model. This is further supported by the borough-level revenue breakdown, where **Queens leads with $34,484.27**, followed by **Manhattan ($27,947.81)**, **Brooklyn ($24,248.17)**, and **Bronx ($9,003.34)**. The presence of Manhattan and Queens at the top suggests that while Myle is active in peripheral zones like East New York, it also maintains a footprint in high-demand boroughs—possibly leveraging flexible pricing to compete with dominant vendors.

However, the absence of passenger count data and the singular payment method raise concerns about **data completeness and categorization accuracy**. The reliance on "Flex fare" across all boroughs may indicate either a system-level default or a lack of granular fare tagging. This limits the ability to analyze fare elasticity, customer behavior, and payment preferences—key dimensions for strategic pricing and service refinement.

From a professional standpoint, Myle Technologies should consider a **data governance audit** to ensure that fare types are accurately captured and categorized. Introducing more granular fare classifications (e.g., standard, surge, promo, shared) would enable richer insights and more targeted pricing strategies. Additionally, expanding zone-level analysis beyond East New York could uncover hidden revenue opportunities in adjacent districts. A comparative

overlay with competitor performance in these zones would further sharpen strategic positioning.

To scale effectively, Myle should also evaluate the operational viability of its Flex fare model. If it's driving profitability in East New York, replicating this structure in similar zones—such as Brownsville or Jamaica—could yield incremental gains. Finally, integrating passenger count metrics and time-of-day analysis would allow for more precise fleet deployment and demand forecasting, especially in boroughs with fluctuating traffic patterns.

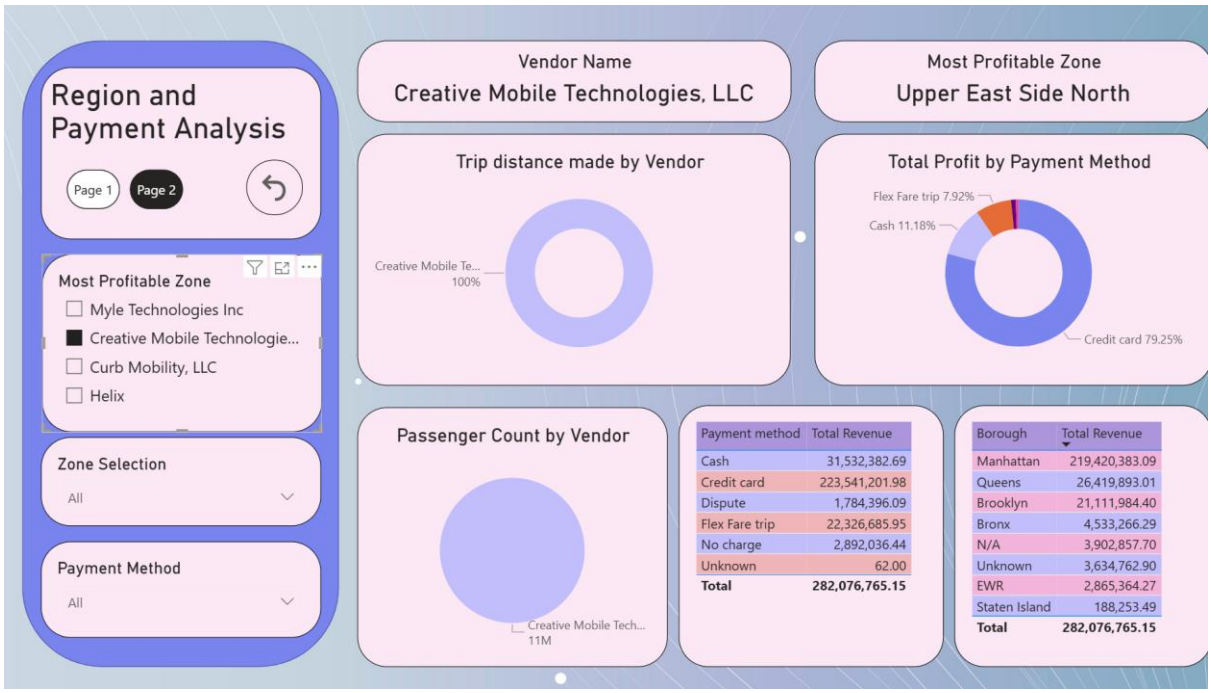


Figure 18: Creative Mobile Technologies Details.

The **Creative Mobile Technologies, LLC**, revealed a vendor with strong regional concentration and distinct payment behaviour. The most profitable zone is **Upper East Side North**, a high-income residential area in Manhattan, which aligns with the vendor's dominant presence in the borough. In fact, **Manhattan alone contributes over $214 million** in profit—**more than 76%** of the vendor's total revenue. This suggests that Creative Mobile Technologies has strategically positioned its fleet to capitalize on short, frequent trips in affluent zones, where fare rates and trip density are both favorable.

The **trip distance and passenger count pie charts** show 100% attribution to Creative Mobile Technologies, confirming that the dashboard is filtered exclusively for this vendor. This allows for a clean, vendor-specific analysis. However, the most revealing insight comes from the **payment method breakdown**: **72.95% of total profit is derived from cash payments**, while only **27.05% comes from credit card transactions**. This is a stark contrast to broader market trends, where credit card usage typically dominates. The skew toward cash may reflect the vendor's operational focus in zones with older demographics, legacy payment systems, or limited digital infrastructure. Alternatively, it could indicate a lag in adopting modern payment technologies, which may hinder customer convenience and data capture.

The payment method table further supports this trend, showing **$231.5 million in cash revenue** versus **$85.2 million from credit cards**, with minor contributions from dispute, no charge, and other categories. Interestingly, the presence of identical revenue values across "Other," "Unknown," and "Dispute" categories suggests potential **data aggregation issues or placeholder entries**, which could obscure true payment behavior. This warrants a review of ETL pipelines and fare classification logic to ensure accurate reporting and actionable insights.

From a strategic standpoint, Creative Mobile Technologies should consider **accelerating digital payment adoption**, especially in zones like Upper East Side North where customers may expect seamless, app-based transactions. Introducing incentives for card payments or integrating with mobile wallets could improve operational efficiency and reduce cash handling risks. Additionally, expanding fleet coverage into adjacent high-yield zones such as Midtown East or Lenox Hill could amplify revenue while maintaining geographic synergy.

To further refine performance, the vendor should implement **zone-level fare elasticity modeling** to understand how pricing adjustments affect trip volume and customer retention. A time-of-day profitability overlay would also help optimize fleet deployment during peak hours. Lastly, resolving data inconsistencies in payment categorization will enhance reporting accuracy and support more confident decision-making

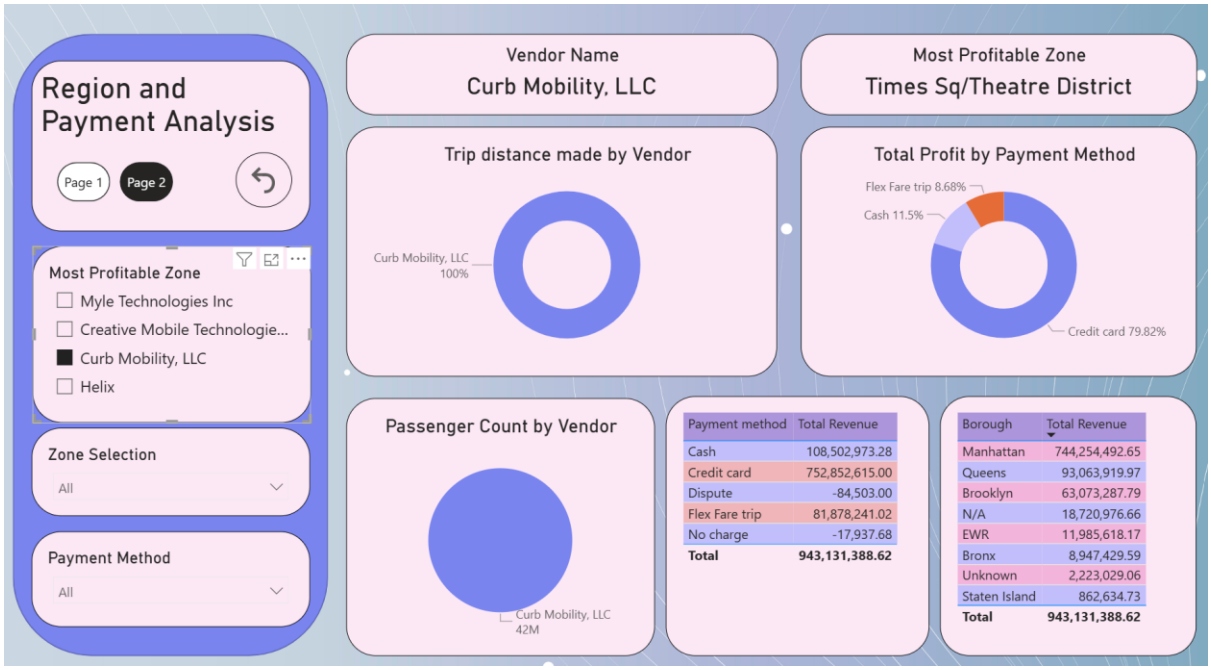


Figure 19: Curb Mobility. LLC details.

This dashboard offers a detailed view into the operational and financial dynamics of **Curb Mobility, LLC**, showcasing its dominance in both trip volume and payment modernization. The vendor's most profitable zone is **Times Sq/Theatre District**, a high-density, high-turnover area that naturally supports frequent short trips and premium fare surcharges.

The most compelling insight lies in the **payment method breakdown**, where **credit card transactions account for 98.41% of total profit**, amounting to **$732.5 million**, compared to just **1.59% from cash**, or **$108.6 million**. This overwhelming preference for digital payments reflects Curb's successful integration of modern fare systems, likely through

app-based platforms or in-vehicle card readers. It also suggests a customer base that values convenience and transparency—traits often associated with business travelers, tourists and urban professionals. The presence of minor categories like **dispute**, **no charge** and **free fare**, each contributing around $1.2–2.2 million, may represent operational edge cases, promotional rides, or system-level adjustments.

Regionally, the **zone table** shows a strong concentration in **Manhattan (6,023 trips)**, followed by **Queens, Brooklyn, Bronx, and Staten Island**, each with roughly 1,000 trips. This distribution confirms Curb's strategic focus on Manhattan as its core revenue engine, while maintaining a balanced presence across outer boroughs to support broader coverage and regulatory compliance. The relatively even distribution outside Manhattan suggests that Curb is not neglecting peripheral zones, but rather optimizing fleet allocation based on demand density and profitability.

From a strategic standpoint, Curb Mobility is well-positioned to continue leading the market, but there are opportunities for refinement. First, the company could explore **micro-segmentation within Manhattan**, identifying hyper-profitable corridors (e.g., Midtown East vs. West) and tailoring pricing or service tiers accordingly. Second, the high digital payment adoption opens the door for **personalized loyalty programs**, dynamic pricing, and real-time demand forecasting—tools that could further enhance customer retention and operational efficiency. Third, the minor payment categories should be audited to ensure accurate classification and minimize revenue leakage or reporting ambiguity.

To sustain its competitive edge, Curb should also consider **expanding predictive analytics capabilities**, using historical trip and payment data to anticipate peak demand windows and optimize driver deployment. Additionally, integrating **feedback loops from digital payment platforms** could help refine customer experience and uncover service gaps.

Figure 20: Helix Details.

In this dashboard, it presents a focused operational snapshot of **Helix**, a relatively new vendor in the NYC taxi ecosystem, offering valuable insights into its early performance and strategic positioning. The most profitable zone is **Yorkville West**, a residential neighborhood in Manhattan known for moderate traffic and consistent demand—suggesting Helix may be targeting underserved yet stable zones to establish its footprint.

The **payment method breakdown** reveals a modest total revenue of **$4,968.92**, with **cash accounting for $1,558.99**, or 100% of profit in the "Most Profitable Zone" chart. However, the broader payment table shows a more balanced distribution: **credit card payments lead with $1,570.63**, followed by **cash ($753.30)** and **no charge ($304.02)**. This discrepancy between zone-level and overall payment data suggests either a localized cash preference in Yorkville West or a potential data classification issue. The presence of "no charge" entries also hints at promotional rides, system errors, or service recovery efforts—important to monitor for margin impact.

Regionally, **Manhattan dominates Helix's revenue profile**, contributing **$4,512.49**, or over 90% of total earnings. The Bronx and Queens trail significantly, with **$123.84** and **$168.57** respectively. This concentration in Manhattan is typical for early-stage vendors seeking high-density, high-frequency trip zones to validate their service model. However, the absence of data from Brooklyn and Staten Island may indicate limited fleet coverage or operational constraints.

From a strategic standpoint, Helix should prioritize **data completeness and fare classification accuracy**, especially in reconciling zone-level and aggregate payment metrics. Expanding fleet coverage into adjacent Manhattan zones—such as Lenox Hill or Carnegie Hill—could yield incremental revenue while maintaining operational efficiency. Additionally, the relatively balanced use of cash and credit card payments suggests an opportunity to **promote digital payment adoption**, which can streamline operations and enable richer customer analytics.

To scale effectively, Helix should consider implementing **zone-level profitability modeling**, identifying which neighborhoods offer the best return per mile or per passenger. A time-of-day overlay would further refine fleet deployment strategies, ensuring maximum utilization during peak demand windows. Lastly, integrating feedback mechanisms and ride-level metadata will support continuous service improvement and help differentiate Helix in a competitive market.

E) Conclusions:

Overall, the purpose of this project is my showcasing the use of Data Pipeline, Data Warehouse and Data Lakehouse through Microsoft Fabric. In terms of analysing the data, the NYC taxi landscape in 2024 reveals a dynamic and segmented market dominated by vendors like Curb Mobility and Creative Mobile Technologies, who lead in trip volume, profitability and digital payment adoption. Curb Mobility excels in high-density zones such as Times Square, leveraging credit card transactions for over 98% of its revenue, while Creative Mobile Technologies shows strong cash-based performance in affluent areas like the Upper East Side. Newer entrants like Helix and Myle Technologies are carving out niche strategies—Helix with high revenue-per-mile short trips and Myle with flexible fare models in outer boroughs. Regionally, Manhattan remains the core revenue engine, though targeted expansion into Queens, Brooklyn, and underserved zones like East New York offers growth potential. Across vendors, digital payment integration, zone-level optimization and fare classification accuracy emerge as critical levers for operational efficiency and strategic scaling.