APPENDIX Uber

Technologies Used:

Kaggle: Source of the original Uber dataset
VS Code: Used as the code editor to run and manage MySQL queries using plugin
MySQL Workbench: Imported and queried cleaned Uber dataset using SQL
Power BI: Dashboard creation and data visualization from cleaned data
Google Cloud Storage: Storage of raw and cleaned datasets in .csv format
Figma: Documentation and UI structuring tool

Data Flow Overview:

Data Collection: Original dataset downloaded from Kaggle (100,000 rows). Uploaded to Google Cloud Storage Bucket for persistence.

Data Import & Cleaning: Imported .csv file into MySQL Workbench using VS Code MySQL extension. Data cleansing and transformations performed via SQL queries:

- Handled null values, incorrect entries.
- Calculated KPIs (revenue per passenger, tip %, etc.)

Post-Processed Data Export: Cleaned dataset exported to a new .csv file. This CSV was stored again in Google Cloud Storage for future use with BigQuery.

Dashboard Creation: Final cleaned .csv file imported into Power BI. Created visual dashboards using DAX measures, calculated columns, and custom visuals.

Documentation & Presentation: Project visuals and workflows documented in Figma for professional presentation.

POWER BI Calculations and Metrics:

Calculated Columns (Table View using DAX)

Trip Category :

```
Formula:
Trip Category =
SWITCH(
  TRUE(),
  'uber'[trip_duration_min]< 5, "Short",
  'uber'[trip_duration_min] >= 5 && 'uber'[trip_duration_min]<= 15, "Medium",
  "Long"
)
```

Description: Categorizes each trip as Short, Medium, or Long based on the trip duration in minutes.

Profit Amount :

```
Formula:
Profit Amount = 'uber'[total_amount] - 'uber'[fare_amount]
```

Description: Calculates the difference between total amount paid and the base fare, representing additional earnings like surcharges and tips.

Tip Behavior :

```
Formula:
Tip Behavior =
SWITCH(
  TRUE(),
  'uber'[tip_percentage] = 0, "No Tip",
  'uber'[tip_percentage] > 0 && 'uber'[tip_percentage] <= 10, "Low Tip",
  'uber'[tip_percentage] > 10 && 'uber'[tip_percentage] <= 20, "Moderate Tip",
  "High Tip"
)
```

Description: Segments rides based on the tip percentage into categories: No Tip, Low Tip, Moderate Tip, and High Tip.

Fare Bracket :

```
Formula:
Fare Bracket =
SWITCH(
  TRUE(),
  'uber'[fare_amount] < 10, "< $10",
  'uber'[fare_amount] >= 10 && 'uber'[fare_amount] < 20, "$10 - $20",
  'uber'[fare_amount] >= 20 && 'uber'[fare_amount] < 40, "$20 - $40",
  "$40+"
)
```

Description: Groups fare amounts into brackets to analyze distribution across fare ranges.

DAX Measures (KPIs and Aggregates)

Total Trips :

```
Formula:
Total Trips = COUNTROWS('uber')
```

Description: Counts the total number of trip records in the dataset.

Total Passengers :

```
Formula:
Total Passengers = SUM('uber'[passenger_count])
```

Description: Calculates the total number of passengers across all trips.

Total Fare Amount :

```
Formula:
Total Fare Amount = SUM('uber'[fare_amount])
```

Description: Calculates the total of all fare amounts across trips.

Total Tip Amount :

```
Formula:
Total Tip Amount = SUM('uber'[tip_amount])
```

Description: Sums up all tip amounts provided by passengers.

Total Trip Distance :

```
Formula:
Total Trip Distance = SUM('uber'[trip_distance])
```

Description: Aggregates the total distance covered by all trips.

Total Duration :

```
Formula:
Total Duration = SUM('uber'[trip_duration_min])
```

Description: Sums the trip durations in minutes for all records.

Total Amount Paid :

```
Formula:
Total Amount Paid = SUM('uber'[total_amount])
```

Description: Calculates the total amount paid by all passengers including fare, tip, tolls, etc.

Avg Fare Amount :

```
Formula:
Avg Fare Amount = AVERAGE('uber'[fare_amount])
```

Description: Calculates the average fare amount per trip.

Avg Tip Amount :

```
Formula:
Avg Tip Amount = AVERAGE('uber'[tip_amount])
```

Description: Computes the average tip amount across all trips.

Avg Tip Percentage :

```
Formula:
Avg Tip Percentage = AVERAGE('uber'[tip_percentage])
```

Description: Computes the average tip percentage based on fare amounts.

Avg Trip Distance :

```
Formula:
Avg Trip Distance = AVERAGE('uber'[trip_distance])
```

Description: Calculates the average distance traveled per trip.

Avg Fare per Mile :

```
Formula:
Avg Fare per Mile = AVERAGE('uber'[fare_per_mile])
```

Description: Calculates how much, on average, is charged per mile.

Avg Fare per Minute :

```
Formula:
Avg Fare per Minute = AVERAGE('uber'[fare_per_minute])
```

Description: Calculates the average cost per minute of trip time.

Tip % of Fare :

```
Formula:
Tip % of Fare = DIVIDE(SUM('uber'[tip_amount]), SUM('uber'[fare_amount]), 0)
```

Description: Shows tip as a percentage of the fare for total trips.

Trips with Tip :

```
Formula:
Trips with Tip =
CALCULATE(COUNTROWS('uber'), 'uber'[tip_amount] > 0)
```

Description: Counts the number of trips where a tip was given.

Trips with Tip :

```
Formula:
% Trips with Tip =
DIVIDE([Trips with Tip],[Total Trips], 0)
```

Description: Calculates the percentage of total trips that included a tip.

Total Fare by Payment Type :

```
Formula:
Total Fare by Payment Type =
CALCULATE(SUM('uber'[fare_amount]), ALLEXCEPT('uber', 'uber'[payment_type_label]))
```

Description: Calculates total fare, grouped by payment type (e.g., cash, card).

Tip % by Payment Type :

```
Formula:
Tip % by Payment Type =
CALCULATE(AVERAGE('uber'[tip_percentage]), ALLEXCEPT('uber', 'uber'[payment_type_label]))
```

Description: Computes average tip percentage grouped by payment type.

Trips by Hour :

```
Formula:
Trips by Hour = COUNTROWS('uber')
```

Description: Counts total number of trips per hour. Best used with hourly time fields in visuals.

Glossary of Terms:

Trip: A single ride or journey taken by a passenger from a pickup location to a drop-off location.

Fare Amount: The base fare charged for a trip, excluding any tips, tolls, or additional fees.

Tip Amount: The extra amount paid by a passenger as a gratuity to the driver.

Total Amount: The full amount charged to the passenger, including fare, tip, tolls, surcharges, and other applicable fees.

Trip Distance: The total distance traveled during a trip, usually measured in miles or kilometers.

Trip Duration: The total time taken for a trip, measured in minutes.

Passenger Count: The number of passengers associated with a single trip.

Vendor ID: A unique identifier representing different taxi or ride-sharing vendors/providers.

Payment Type: The mode of payment used for a trip (e.g., cash, credit card, digital wallet).

Fare per Mile: A calculated metric representing the average fare amount paid per mile of distance traveled.

Fare per Minute: A calculated metric representing the average fare paid per minute of trip duration.

Tip Percentage: The tip amount as a percentage of the fare amount.

Pickup Hour: The hour of the day when the trip began, used for time-based analysis.