



TRANSIT OPERATION ANALYSIS - POWER BI REPORT

This Power BI Report was developed to provide a clear weekly view of on-time performance, Total Rides, completed rides e.t.c I built a clean hybrid star-schema data model centered on a fact table of rides with dimensions for date, passenger, location, shift and vehicle. There are three main pages in this report namely:

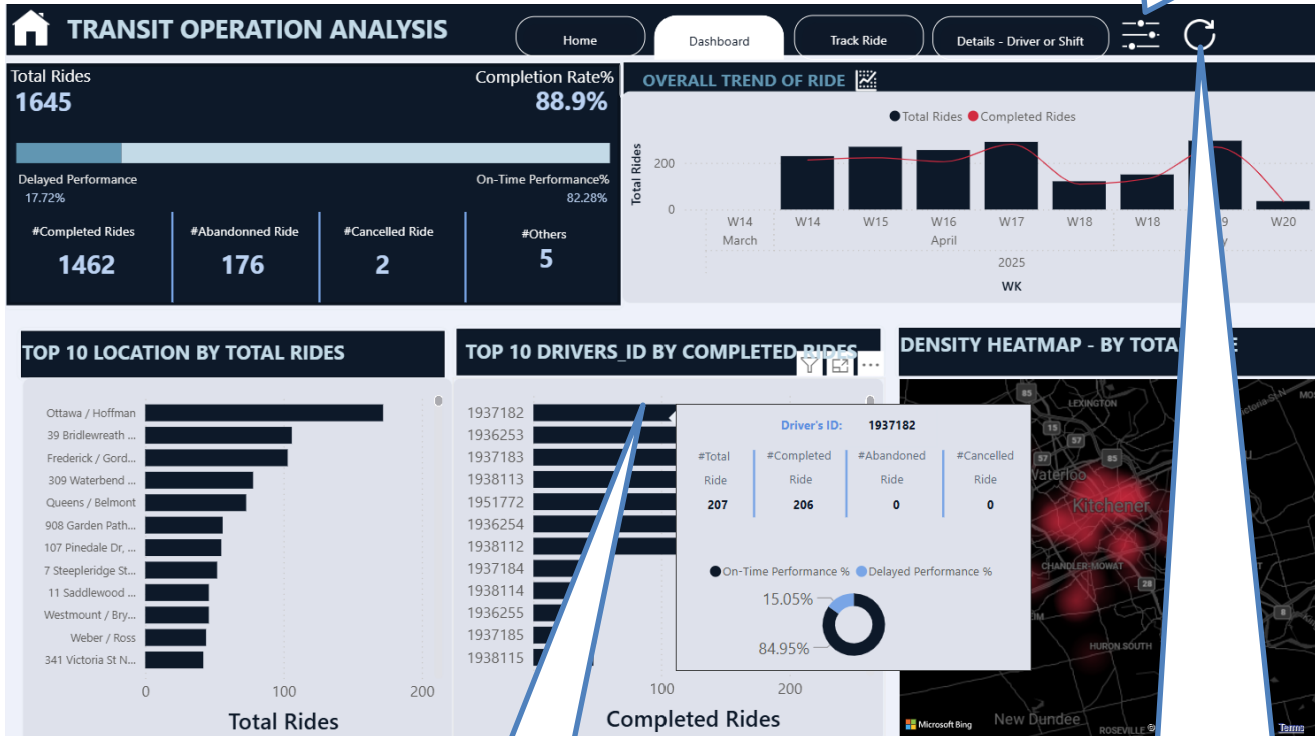
- Overview
 - Trip Summary Insight and
 - Shift Details
-

The main report page "Overview" surfaces a small set of high impact KPI at the top: Total Rides, Completion Rate, On-Time performance%, Abandoned rides, Cancelled Rides and Average Trips. These metrics were further contextualised with a trend over-time visual (Total Rides and Completed Ride by week with drill down capability).

From the sample period the network shows a high ride completion rate (~ 89%) and a solid on-time delivery (~82%), but also a notable share of abandoned rides.

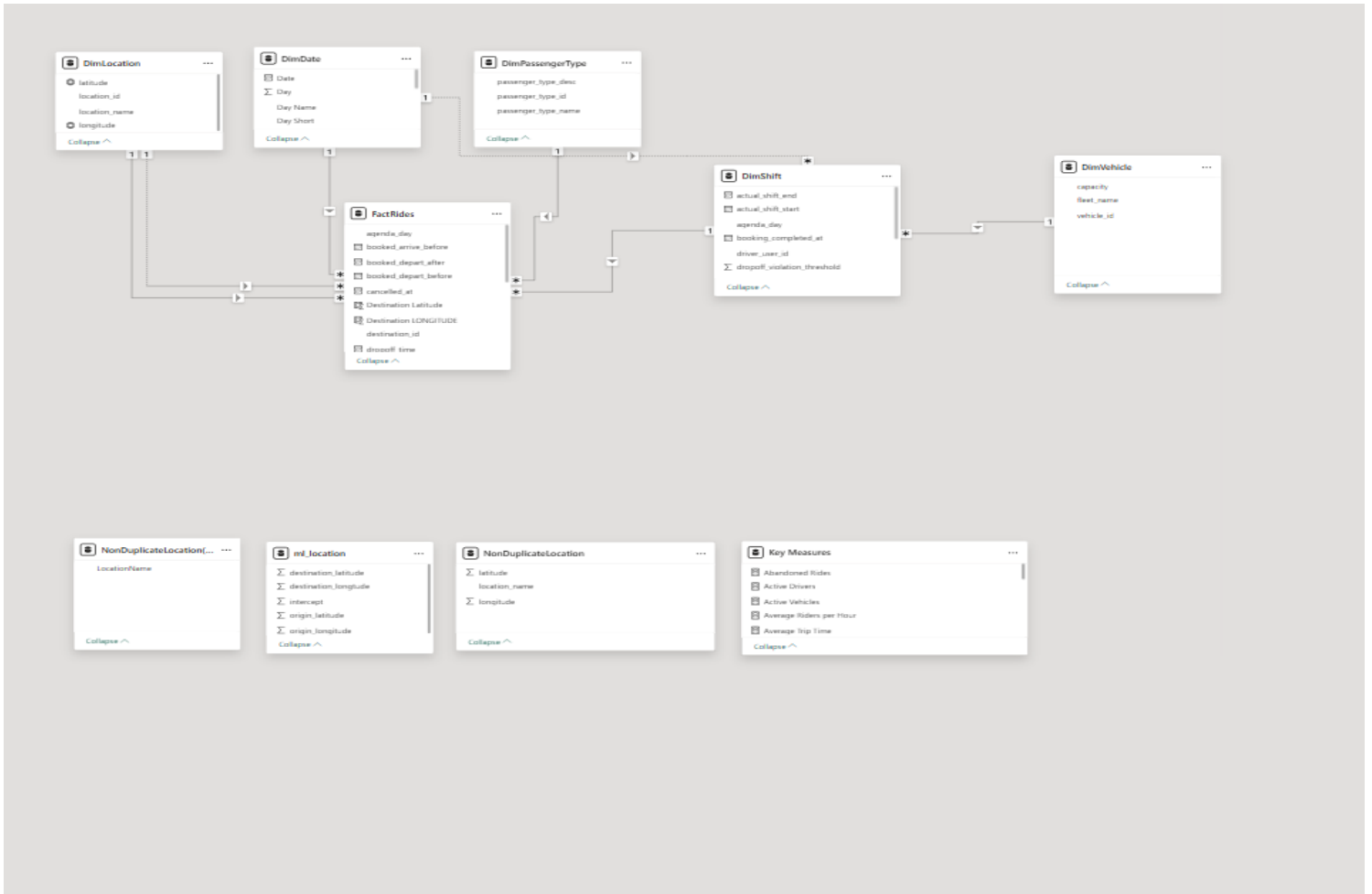
To incorporate interactivity for the users, the report utilizes bookmarks and drilldown capabilities to enrich user experience. Enriched visual tooltips were not left out as they provided additional insight into some of the visual e.g. the **Top 10 Driver_ID by completed rides**

Additional pane was added using bookmarks to provide more filter capability for users



Button to clear all slicers in this current page

Fill Free right click and drill down for more detailed analysis into each driver.



Our data model is a Hybrid Star-Snowflake Schema but leans 90% towards a Star Schema.

Manage relationships

×

+ New relationship

⚡ Autodetect

Edit

Delete

Filter

<input type="checkbox"/> From: table (column) ↑	Relationship	To: table (column)	Status
<input type="checkbox"/> DimShift (agenda_day)	* — ◀ — 1	<input checked="" type="checkbox"/> DimDate (Date)	Inactive ...
<input type="checkbox"/> DimShift (vehicle_id)	* — ◀ — 1	DimVehicle (vehicle_id)	Active ...
<input type="checkbox"/> FactRides (agenda_day)	* — ◀ — 1	<input checked="" type="checkbox"/> DimDate (Date)	Active ...
<input type="checkbox"/> FactRides (destination_id)	* — ◀ — 1	DimLocation (location_id)	Inactive ...
<input type="checkbox"/> FactRides (origin_id)	* — ◀ — 1	DimLocation (location_id)	Active ...
<input type="checkbox"/> FactRides (passenger_type_ids)	* — ◀ — 1	DimPassengerType (passenger...	Active ...
<input type="checkbox"/> FactRides (shift_id)	* — ◀ — 1	DimShift (shift_id)	Active ...

This Section provides insight into some of the DAX measures used in the transit operation Dashboard. They were provided to give a clear view of the thought behind some the measures created.

1. Total Rides

The total number of ride records in the system regardless of status.

This metric represents overall demand for transit services during the selected time period. It reflects all ride requests across the network.

Calculation Logic:

This measure simply counts the number of rows in the FactRides table:

- Includes completed, canceled, abandoned, no-shows, and in-progress rides.

2. Completed Rides

The number of rides that were successfully completed. This represents rides where service was fully delivered.

Calculation Logic:

The measure filters rides where:

ride_status = "COMPLETED"

3. On-Time Performance (%)

The percentage of completed rides that met both pickup and dropoff time promises.

.

A ride is counted as On-Time if:

Pickup occurs between the promised pickup window:

$\text{pickup_time} \leq \text{booked_depart_before}$

Dropoff occurs before the promised arrival time

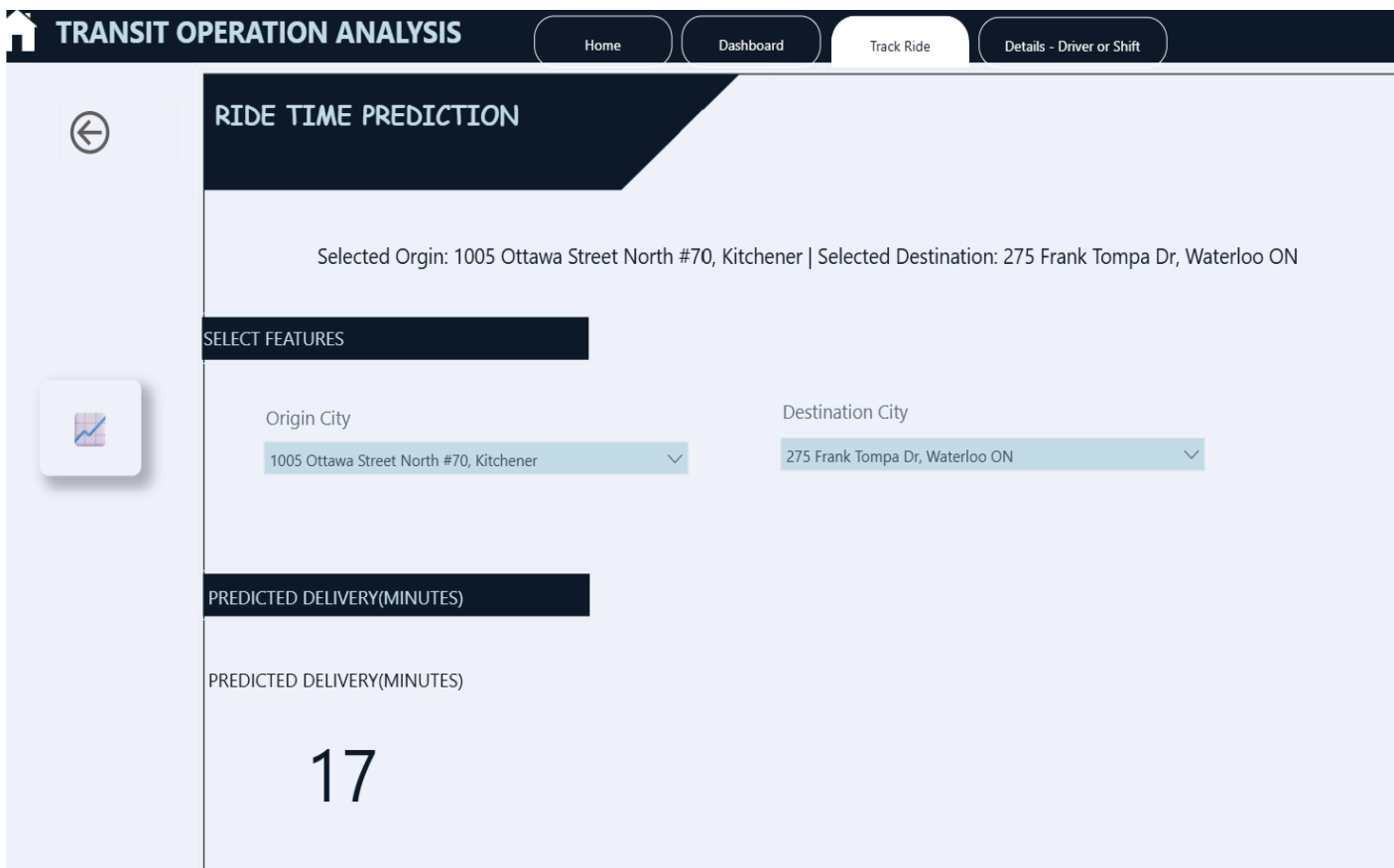
$\text{dropoff_time} \leq \text{booked_arrive_before}$

Formula Summary:

$$\text{On-Time Performance \%} = \frac{\text{On-Time Completed Rides}}{\text{Total Completed Rides}}$$

Additional Interesting Capability of the BI Report

This Power BI report incorporates a Machine Learning regression model to provide intelligent ride time predictions. Users can select an origin and destination location, and the model dynamically calculates the estimated trip duration based on historical ride data.



TRANSIT OPERATION ANALYSIS

Home Dashboard **Track Ride** Details - Driver or Shift

RIDE TIME PREDICTION

Selected Orgin: 1005 Ottawa Street North #70, Kitchener | Selected Destination: 275 Frank Tompa Dr, Waterloo ON

SELECT FEATURES

Origin City
1005 Ottawa Street North #70, Kitchener

Destination City
275 Frank Tompa Dr, Waterloo ON

PREDICTED DELIVERY(MINUTES)

PREDICTED DELIVERY(MINUTES)

17

Technical Implementation

- Regression model built in Python (scikit-learn) using historical ride data with features including origin/destination coordinates, distance, and time patterns
- Model trained on completed rides dataset to learn relationships between geographic coordinates and actual trip durations
- Prediction results exported and stored in Delta table format for optimized querying and data versioning
- Delta table connected to Power BI as data source and integrated into the semantic model alongside operational tables
- Interactive slicers enable users to select origin/ Destination location triggering Dax MEASURES to compute the average time for that trip.



Report Refresh

To operationalize this report for weekly or daily reviews this report would be published to the power BI service after which we

- we can configure refresh for daily or hourly e.t.c.
- I would also ensure all CSV file are connected via a Gateway
- Set refresh failure notification for proactive monitoring

If the data asset grows 100x, I would recommend

1. Implement **Incremental refreshes** eg on theFactRide table

2. Move to a lakehouse + Delta architecture(Medallion model).

Using Fabrics' recommended Bronze –Silver –Gold architecture would reduce dataset size and significantly speed up report.

Sharing Securely

- Publish report to Power BI workspace and utilize Power BI app to bundle all dashboards and reports
- Assign role based permissions: eg Admin to provide full access, Drivers/Operators with View permissions
- Implement RLS to restrict data at the role level:
eg Driver -based filtering (so that drivers only see their own ride)