

# Namma Yatri Data Analysis: Technical Methodology & Implementation

A Deep Dive into Data Preparation, Modeling, and Analytical Techniques

Group Members: R.Gayathri, Souvik Biswas, Savya Sharma



# Objective & Agenda

**Objective:** To provide a comprehensive overview of the technical methodology applied to the Namma Yatri dataset, detailing data acquisition, preparation, modeling, analytical techniques, and key technical insights.

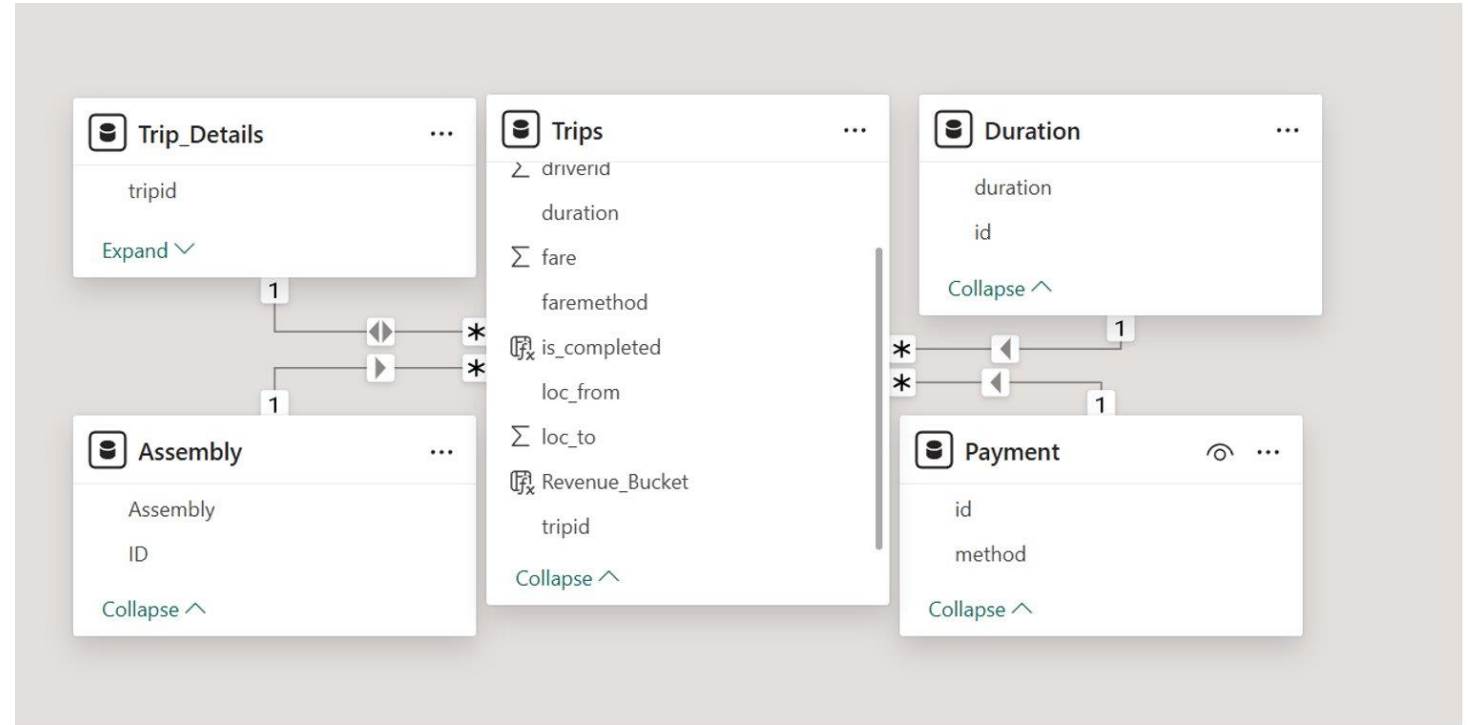
## Agenda:

- Project Overview & Data Sources
- Data Acquisition & Preparation Pipeline
- Data Modeling & Relationships
- Exploratory Data Analysis Techniques
- Key Technical Findings & Visualizations
- Challenges Encountered & Solutions
- Conclusion



# Project Overview & Data Sources

- Analysis of Namma Yatri ride-hailing data.
- **5 Tables from nammayatri.xlsx:**
- Trips, Trip\_Details (Core Trip Info)
- Assembly (Zone Names)
- Duration (Hourly Time Bins)
- Payment (Payment Methods)



# Data Acquisition & Preparation Pipeline

- **Acquisition:** Direct import from `nammayatri.xlsx` into Power BI Desktop.

## Power Query Steps:

- **Data Type Standardization:** Crucial for `tripid`, `fare`, `duration` (Whole/Decimal Number).
- **Consistency Checks:** Ensured clean data for relationships.

Queries [5] <

fx = Table.TransformColumnTypes("#Promoted Headers",{"tripid", Int64.Type}, {"faremethod", Int64.Type}, {"fare", type number}, {"loc\_from",

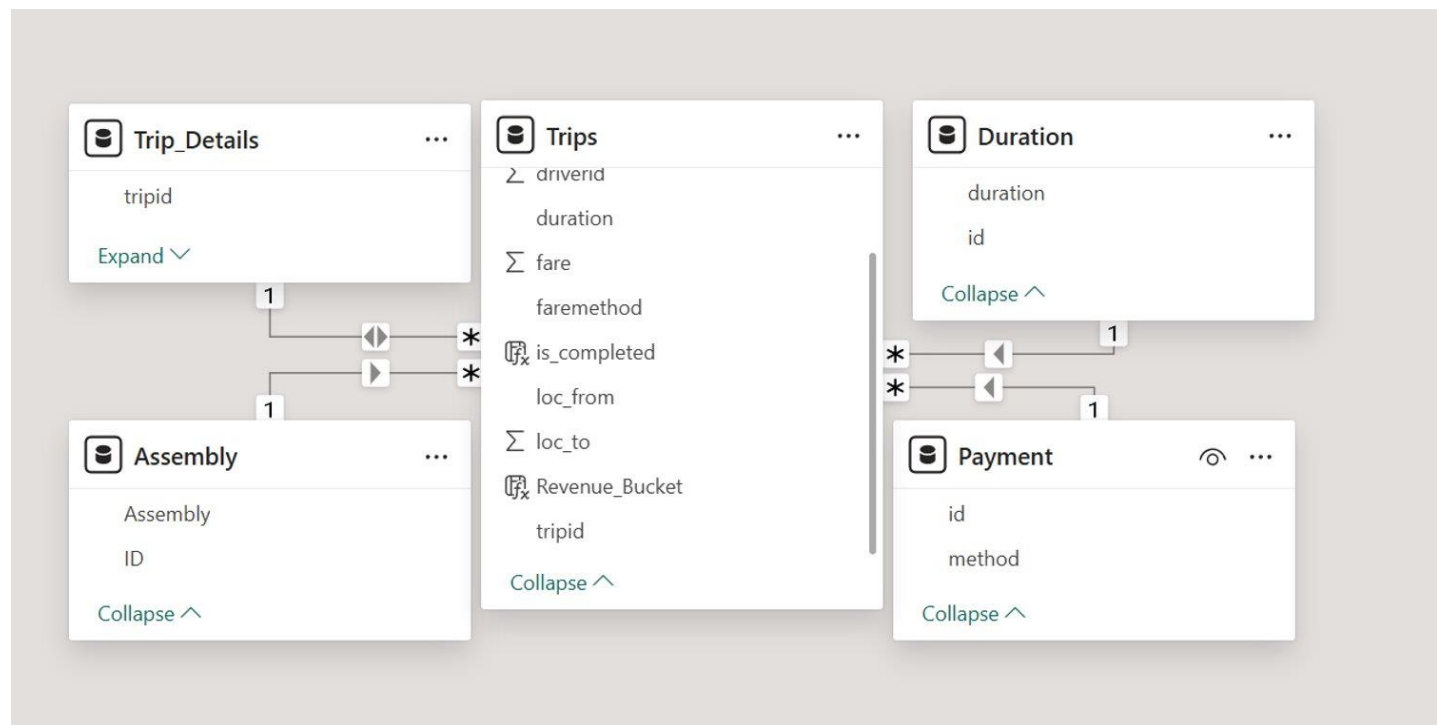
	123 tripid	123 faremethod	1.2 fare	123 loc_from	123 loc_to	123 driverid	123 custid
	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%	Valid 100% Error 0% Empty 0%
1	1	2	776	16	8	7	
2	2	2	1479	24	33	1	
3	3	4	152	18	32	6	
4	4	3	153	28	14	28	
5	5	2	366	19	27	27	
6	6	1	963	37	2	14	
7	7	3	581	22	16	15	
8	8	1	918	28	3	5	
9	9	4	700	25	27	4	

# Data Modeling & Relationships

A robust star schema-like data model was constructed in Power BI's Model View to facilitate efficient cross-table analysis and filtering. This structure allows dimension tables (Assembly, Duration, Payment) to filter the fact tables (Trips, Trip\_Details).

## Key Relationships Established:

- Trips [tripid] (**Many**) <--> (**One**) Trip\_Details [tripid]
  - Type: Many-to-One (assuming Trip\_Details [tripid] is unique per trip, and Trips might have multiple records if a trip is defined differently or linked to another context.) If tripid is truly unique in both, it's 1:1.
  - Cross-filter Direction: **Both** (to allow filters to flow in either direction, crucial for complex calculations linking Duration to Trip\_Details via Trips).
- Trips [loc\_from] (**Many**) --> (**One**) Assembly [Assembly\_ID]
  - Type: Many-to-One.
  - Cross-filter Direction: Single.
- Trips [duration] (**Many**) --> (**One**) Duration [id]
  - Type: Many-to-One.
  - Cross-filter Direction: Single.
- Trips [faremethod] (**Many**) --> (**One**) Payment [id]
  - Type: Many-to-One.
  - Cross-filter Direction: Single.
  - Ensuring correct cardinality and filter direction was paramount for accurate measure calculations.



# Exploratory Data Analysis Techniques

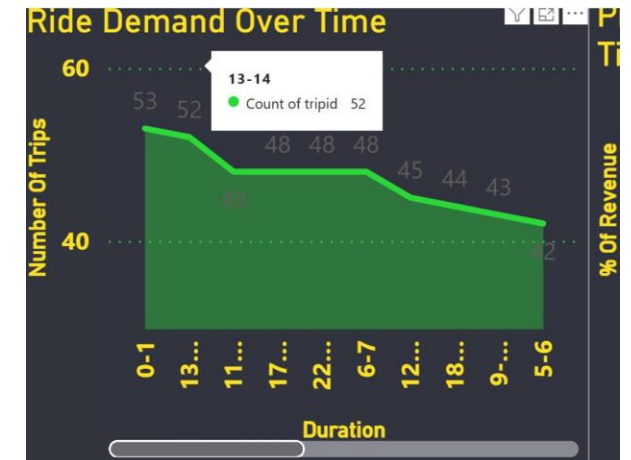
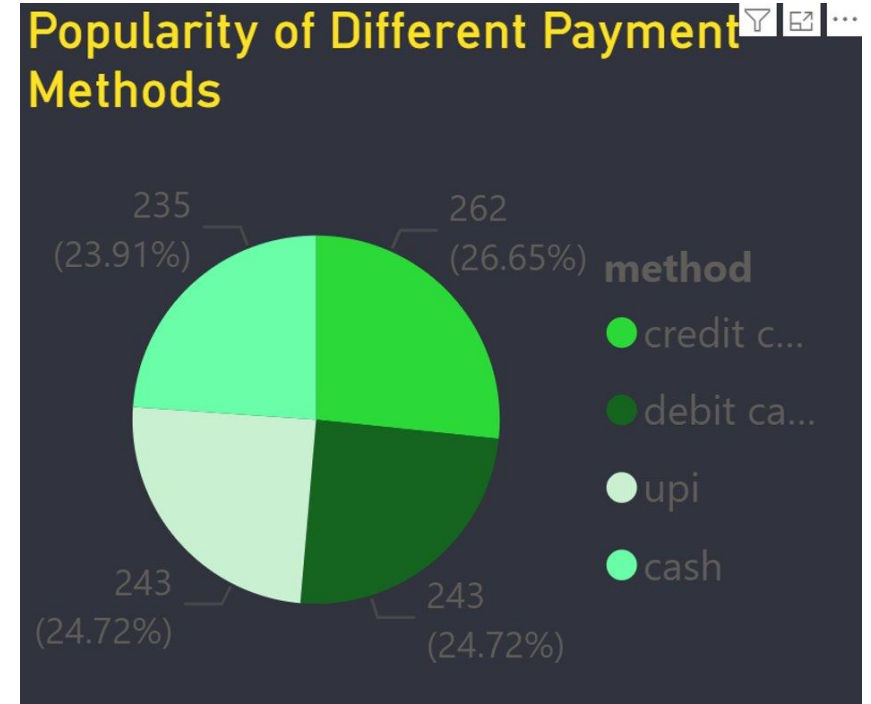
---

- **Variable Classification:**

- Categorical: Zone, Hour, Payment Method, IDs.
- Numerical: Fare, Distance, Trip Counts, Binary Flags.

- **Methods Applied:**

- Time-Series Trends (Line charts for Demand/Revenue).
- Categorical Distributions (Pie/Bar charts for Payment, Cancellation).
- Ranking & Top N Analysis (Bar charts for Zones).
- Ratio/Percentage Calculations (DAX measures for conversion/cancellations).



# Key Technical Findings & Visualizations

- Visuals served as critical tools for **technical validation** and **uncovering data nuances**.
- **Data Quality:** Identified and confirmed data consistency through visual inspection.
- **Model Validation:** Verified correct relationship propagation and filter contexts across tables.
- **DAX Measures:** Confirmed accuracy of complex measures (e.g., cancellation rates, completion percentages) through visual outputs.
- **Unexpected Patterns:** Highlighted intrinsic data behaviors, such as the consistent 83.4% completion rate for trips with quote searches, directly from the data.

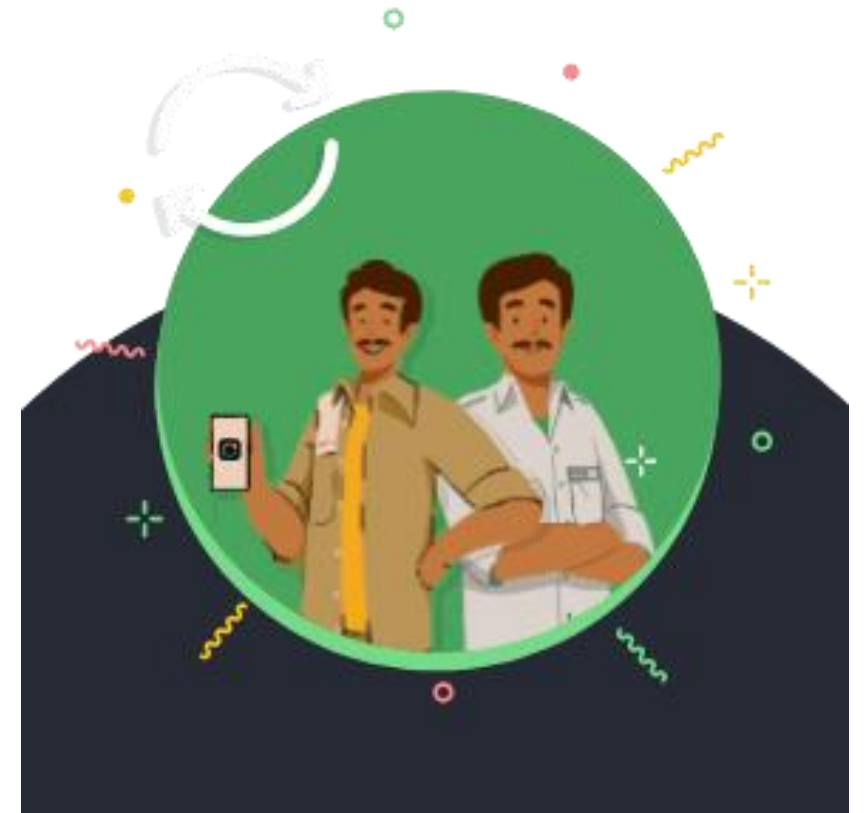




# Technical Challenges & Solutions

---

- **Data Consistency:**
  - *Challenge:* Ensuring uniform data types for relationships.
  - *Solution:* Rigorous type conversion in Power Query.
- **2. Data Model Filter Context:**
  - *Challenge:* Accurate percentage calculation across related tables.
  - *Solution:* Precise DAX measures and appropriate relationship settings (e.g., 'Both' filter direction).
- **3. Power BI UI Issue (Parameter):**
  - *Challenge:* Drag-and-drop functionality issue for parameter creation.
  - *Solution:* Acknowledged as an environmental glitch; focused on conceptual understanding of parameters.





# Conclusion

---

- Successfully established a clean and robust Power BI data model.
- Executed comprehensive EDA with varied techniques and DAX.
- Demonstrated ability to uncover subtle data patterns through precise analytical framework.
- Resulting dashboards are technically sound and ready for business interpretation.



**namma  
yatri**

# Appendix: Key DAX Measures

- Total Trips = COUNTROWS(Trips)
- Total Revenue = SUM(Trips[fare])
- Pct Successful Rides = DIVIDE(CALCULATE(COUNTROWS('Trip Details'), 'Trip Details'[end\_ride] = 1), COUNTROWS('Trip Details'))
- Pct Mutually Cancelled = DIVIDE(CALCULATE(COUNTROWS('Trip Details'), 'Trip Details'[driver\_not\_cancelled] = 0, 'Trip Details'[customer\_not\_cancelled] = 0), COUNTROWS('Trip Details'))
- Pct Completed After Quote Search =  
DIVIDE(CALCULATE(COUNTROWS('Trip Details'), 'Trip Details'[searches\_for\_quotes] = 1, 'Trip Details'[end\_ride] = 1), CALCULATE(COUNTROWS('Trip Details'), 'Trip Details'[searches\_for\_quotes] = 1))