

A decorative graphic on the left side of the slide, consisting of a network of light blue lines and small circles, resembling a circuit board or data flow diagram.

UBER ETL PIPELINE DATA ENGINEERING PROJECT



By Prince



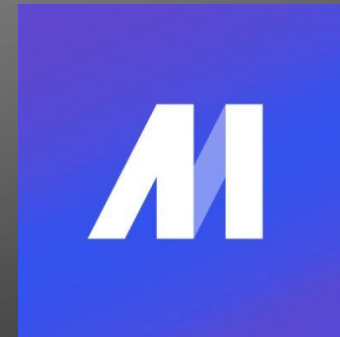
TABLE OF CONTENTS

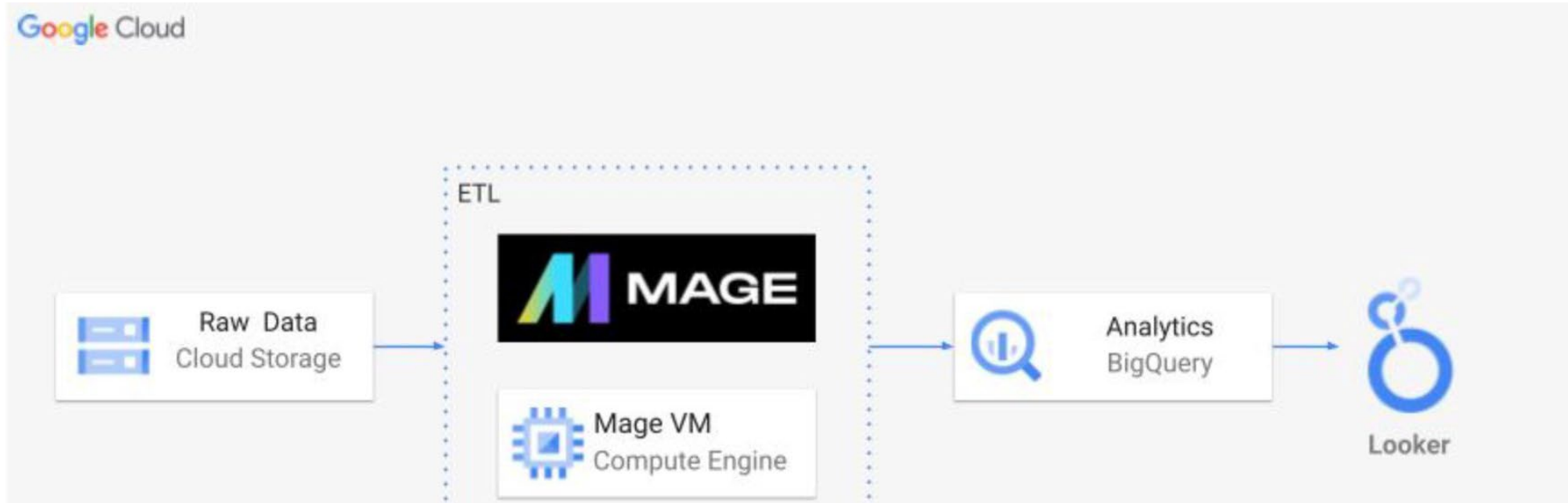
- About the dataset and Tech Stack
- Data Architecture
- Data Dimensional Modeling
- Data Transformation (Mage)
- BigQuery Analytics
- Dashboard
- Key Insights and Findings

ABOUT THE DATASET

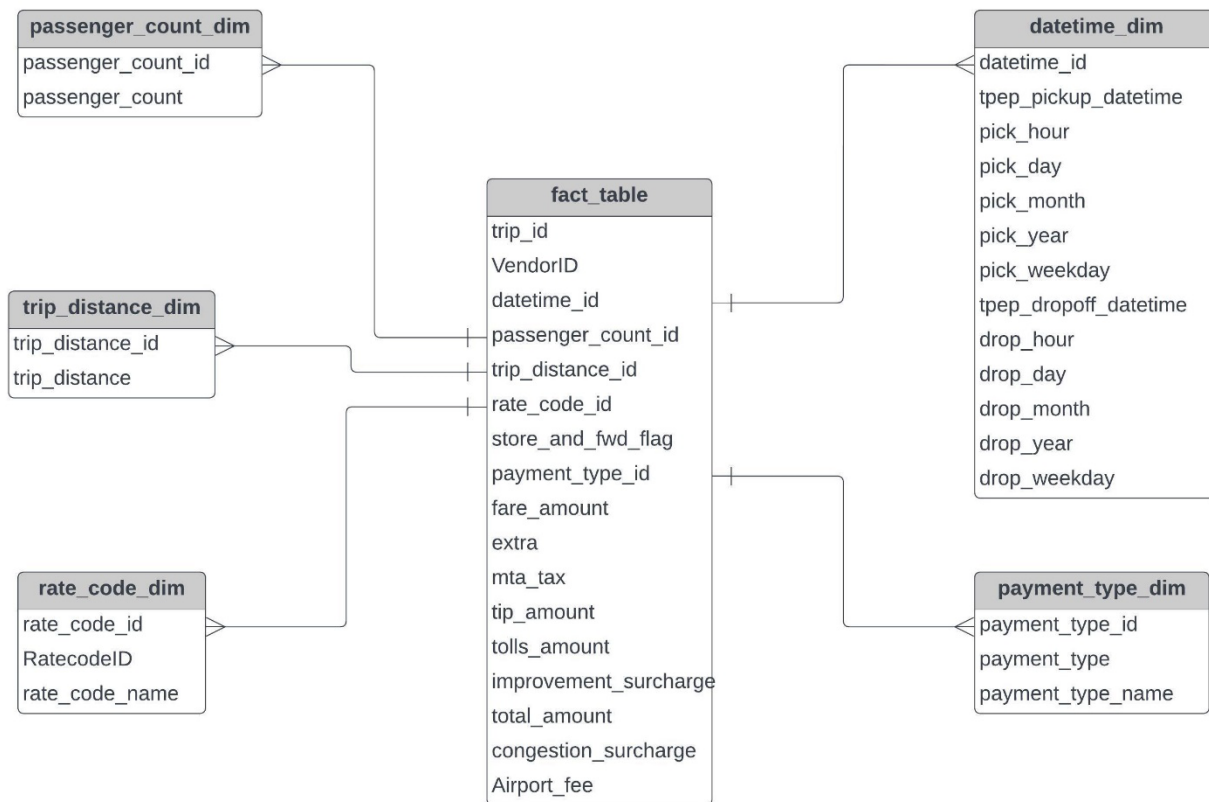
- This Dataset is of the uber company for the month of March 2023.
- There are 100,000+ rows in the dataset.
- It mainly contains information about the rides

TECH STACK





DATA ARCHITECTURE DIAGRAM



DATA DIMENSIONAL MODELING

DATA ETL (MAGE)

The screenshot displays the Mage Data ETL interface. On the left, a file explorer shows the project structure under 'uber-de-project', including folders like 'charts', 'custom', 'data_exporters', 'data_loaders', 'dbt', 'extensions', and 'pipelines'. The 'hidden_wind' pipeline is selected. The main editor shows three steps in the pipeline:

- DATA LOADER** (load_data): (27 lines collapsed)
- TRANSFORMER** (data_transformation): (154 lines collapsed)
- DATA EXPORTER** (load_bigquery): (33 lines collapsed)

Below the steps, there are buttons for adding components: Data loader, Transformer, Data exporter, DBT model, Custom, Scratchpad, Sensor, and Markdown. On the right, a visual pipeline diagram shows the flow from 'load_data' to 'data_transformation' to 'load_bigquery'. Each step is represented by a box with an icon and a checkmark, indicating successful execution.

Loaded the data from Google Cloud Storage

Python script to transform the data

Load the transformed table to Google Big Query

BIGQUERY ANALYTICS

Creating the final table with useful columns from the dataset for visualizations.

```
1 CREATE OR REPLACE TABLE `ethereal-reef-393116.uber_de_project.uber_analytics` AS
2 (
3 SELECT
4 f.trip_id,
5 f.vendor_id,
6 d.tpep_pickup_datetime,
7 d.tpep_dropoff_datetime,
8 p.passenger_count,
9 t.trip_distance,
10 r.rate_code_name,
11 pay.payment_type_name,
12 f.fare_amount,
13 f.extra,
14 f.mta_tax,
15 f.tip_amount,
16 f.tolls_amount,
17 f.improvement_surcharge,
18 f.total_amount
19 FROM
20
21 `ethereal-reef-393116.uber_de_project.fact_table` f
22 JOIN `ethereal-reef-393116.uber_de_project.datetime_dim` d ON f.datetime_id=d.datetime_id
23 JOIN `ethereal-reef-393116.uber_de_project.passenger_count_dim` p ON p.passenger_count_id=f.passenger_count_id
24 JOIN `ethereal-reef-393116.uber_de_project.trip_distance_dim` t ON t.trip_distance_id=f.trip_distance_id
25 JOIN `ethereal-reef-393116.uber_de_project.rate_code_dim` r ON r.rate_code_id=f.rate_code_id
26 JOIN `ethereal-reef-393116.uber_de_project.payment_type_dim` pay ON pay.payment_type_id=f.payment_type_id)
27 ;
28
```

EXECUTIVE SUMMARY

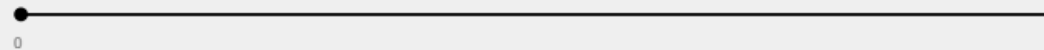
Uber

FILTERS

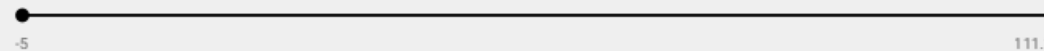
PAYMENT TYPE ▾

RATE TYPE ▾

PASSENGER COUNT



TIP AMOUNT



SUMMARY

TOTAL AMOUNT
\$3.80M

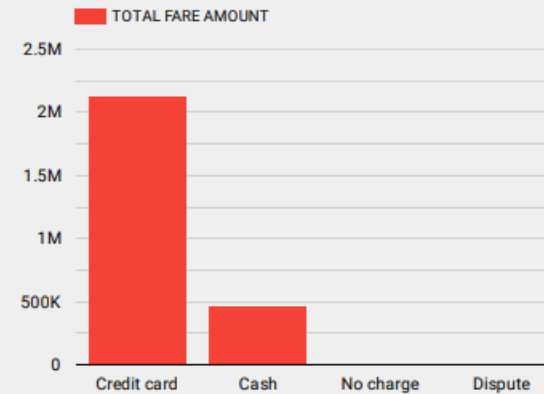
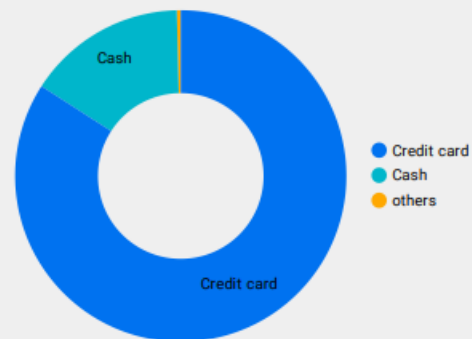
RECORD COUNT
134.6K

AVG. TRIP DIST.
3.5

AVG. FARE AMT.
19.3

AVG. TIP AMT.
3.5

CHARTS



[LINK](#)

