UBER ETL PIPELINE DATA ENGINEERING PROJECT

By Prince

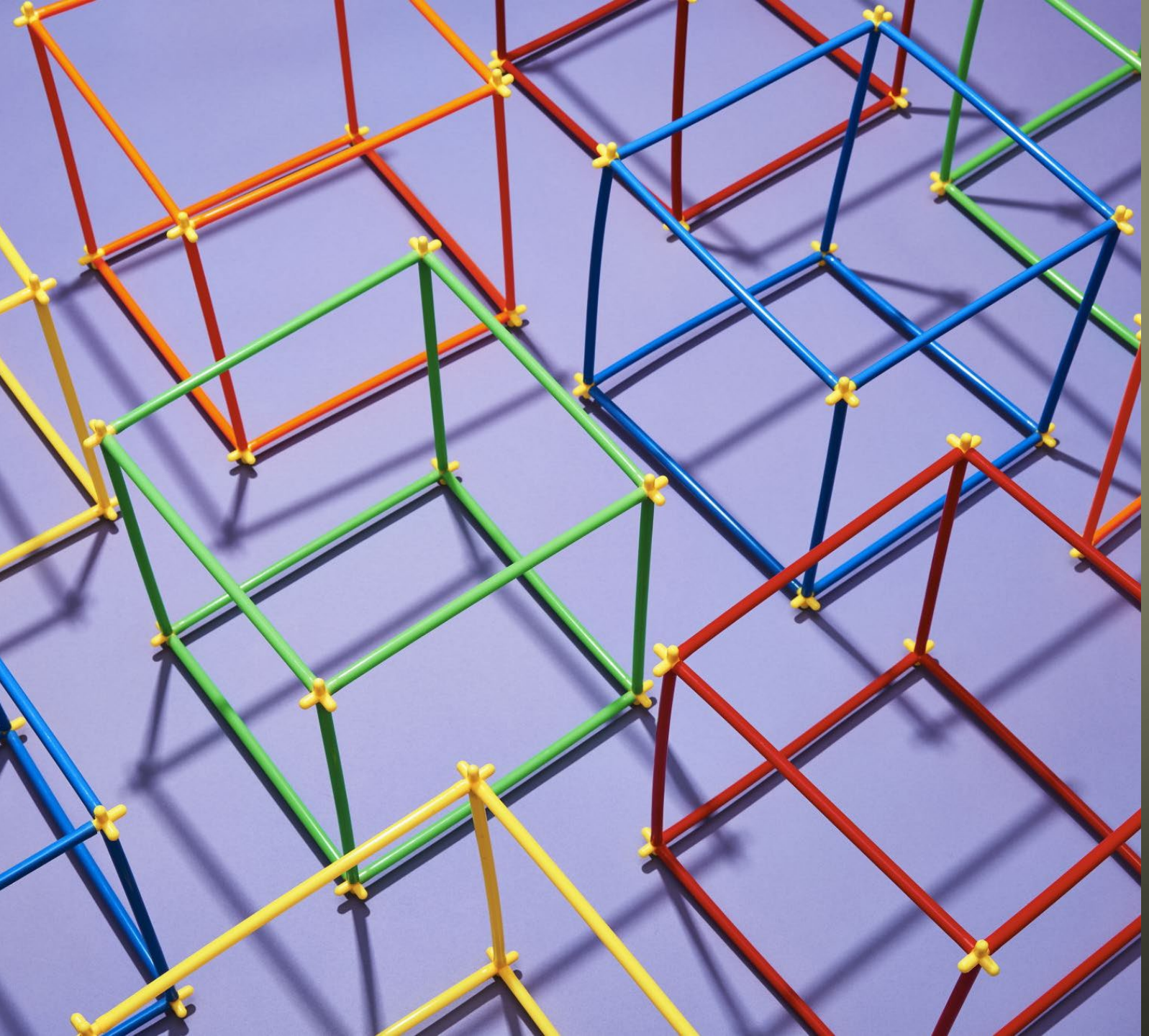
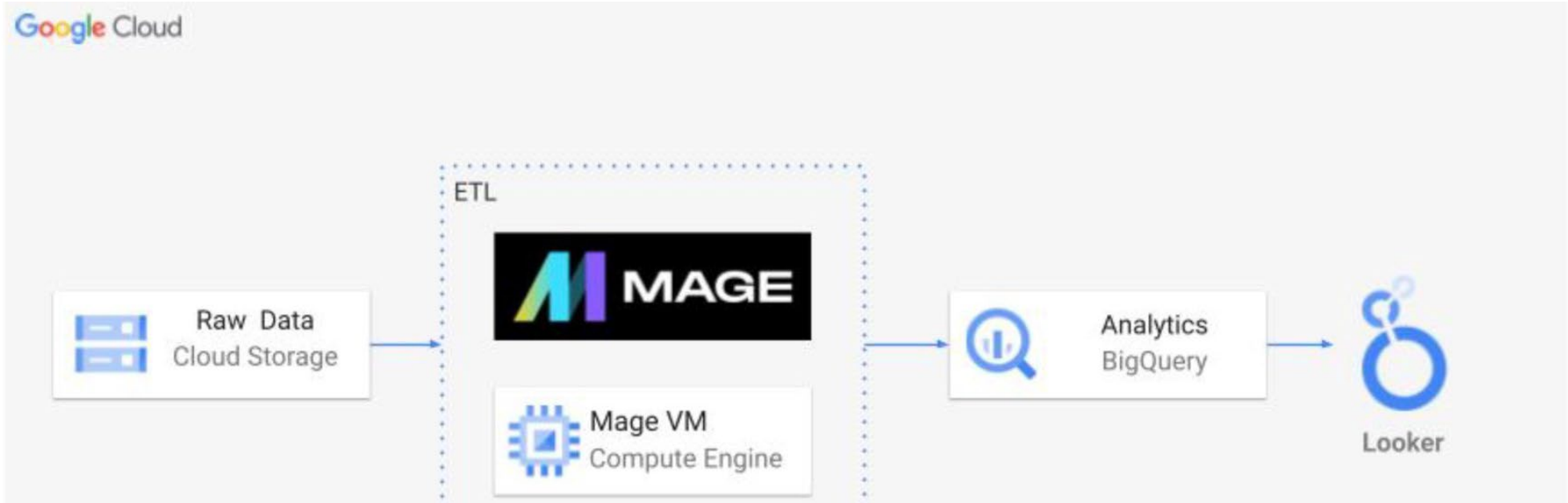


TABLE OF CONTENTS

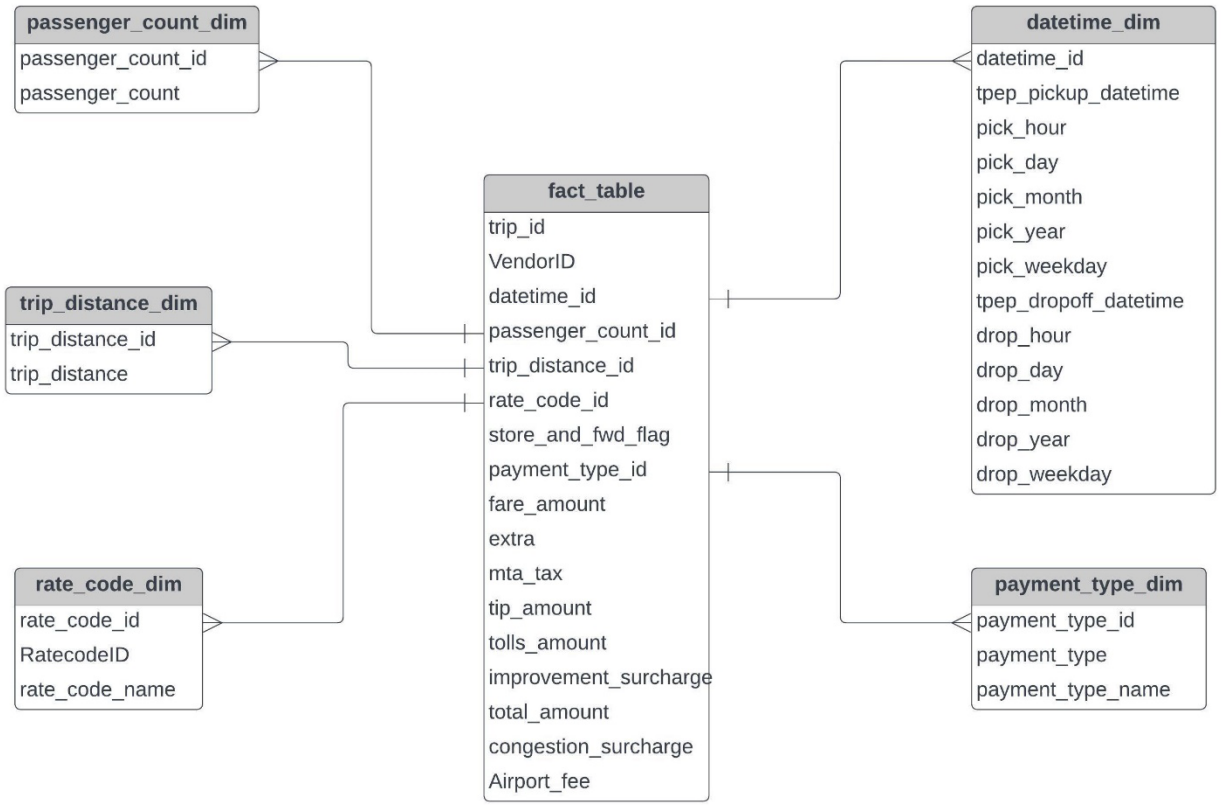
ABOUT THE DATASET

- 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

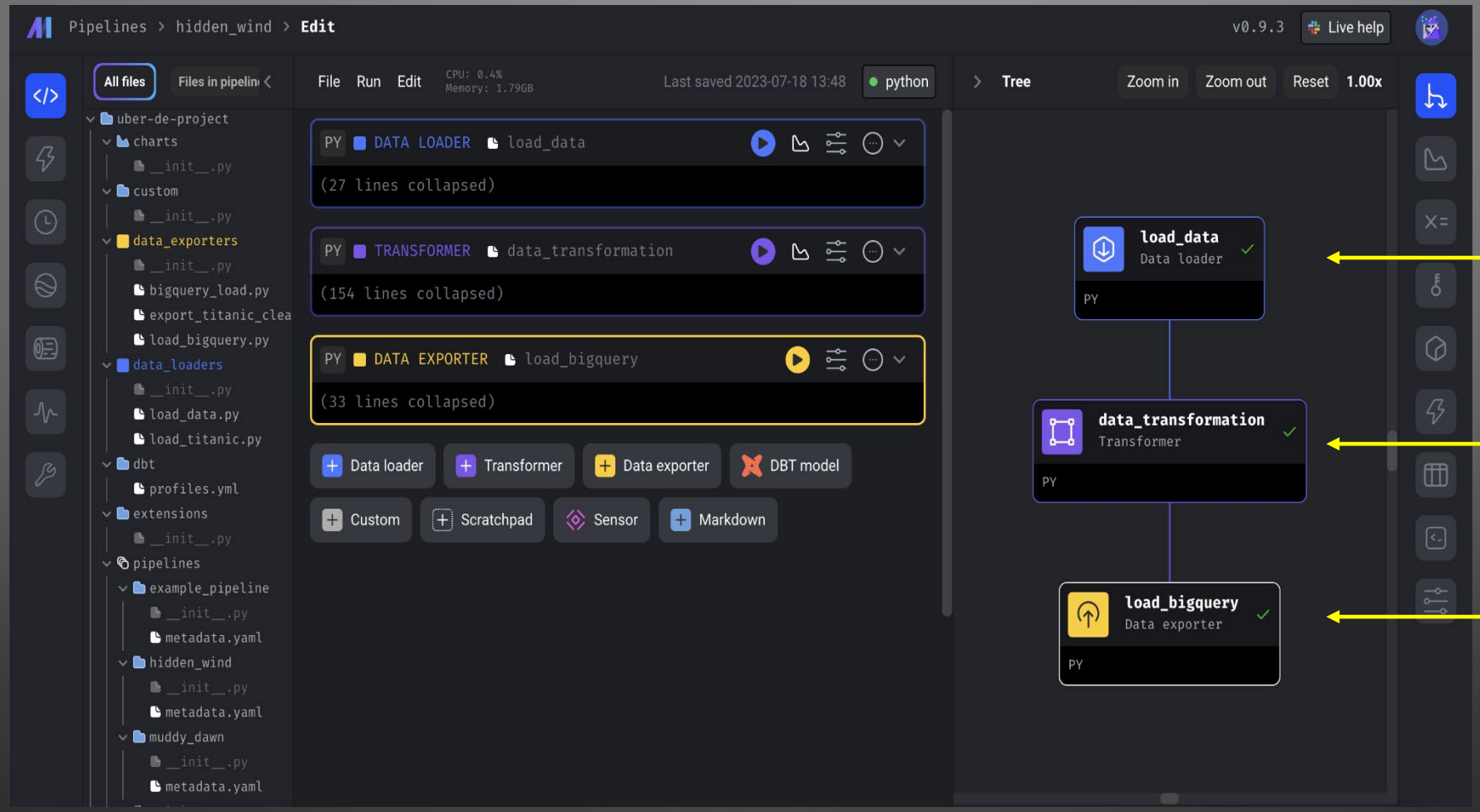


DATA ARCHITECTURE DIAGRAM



DATA DIMENSIONAL MODELING

DATA ETL (MAGE)



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.

```sql
CREATE OR REPLACE TABLE `ethereal-reef-393116.uber_de_project.uber_analytics` AS
(
SELECT
f.trip_id,
f.VendorID,
d.tpep_pickup_datetime,
d.tpep_dropoff_datetime,
p.passenger_count,
t.trip_distance,
r.rate_code_name,
pay.payment_type_name,
f.fare_amount,
f.extra,
f.mta_tax,
f.tip_amount,
f.tolls_amount,
f.improvement_surcharge,
f.total_amount
FROM

`ethereal-reef-393116.uber_de_project.fact_table` f
JOIN `ethereal-reef-393116.uber_de_project.datetime_dim` d  ON f.datetime_id=d.datetime_id
JOIN `ethereal-reef-393116.uber_de_project.passenger_count_dim` p  ON p.passenger_count_id=f.passenger_count_id
JOIN `ethereal-reef-393116.uber_de_project.trip_distance_dim` t  ON t.trip_distance_id=f.trip_distance_id
JOIN `ethereal-reef-393116.uber_de_project.rate_code_dim` r ON r.rate_code_id=f.rate_code_id
JOIN `ethereal-reef-393116.uber_de_project.payment_type_dim` pay ON pay.payment_type_id=f.payment_type_id)
;
```

EXECUTIVE SUMMARY

Uber

FILTERS

PAYMENT TYPE

RATE TYPE

PASSENGER COUNT

0 — 9

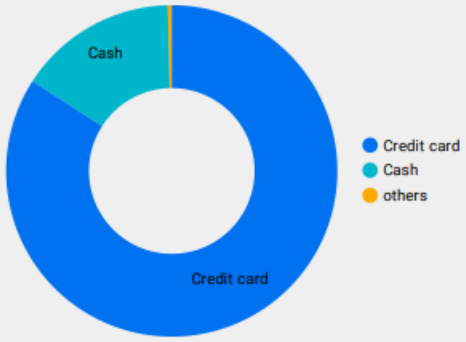
TIP AMOUNT

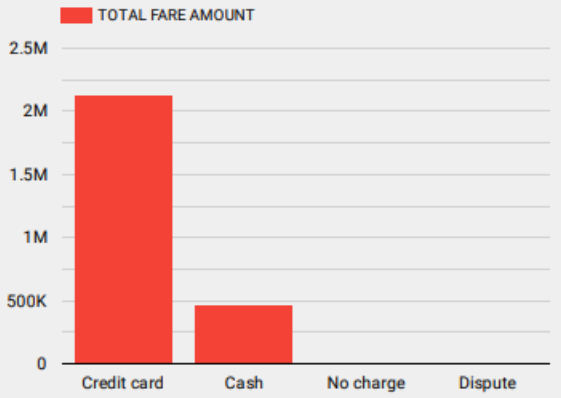
-5 — 111.11

SUMMARY

TOTAL AMOUNT	RECORD COUNT	AVG. TRIP DIST.	AVG. FARE AMT.	AVG. TIP AMT.
$3.80M	134.6K	3.5	19.3	3.5

CHARTS





LINK