**Introduction:**

In the fast-paced world of ride-sharing and transportation, Uber has revolutionized the way people move around cities. Behind this seamless experience lies a complex and robust data infrastructure that enables Uber to provide efficient, reliable, and personalized services to millions of users worldwide. The "Uber Data Analytics Project" is an ambitious undertaking aimed at enhancing Uber's data processing capabilities and optimizing their decision-making processes. This comprehensive data pipeline aims to collect, process, store, and analyze vast amounts of data generated by Uber's operations, ensuring that the company can extract valuable insights to improve their services continually.

In this project we will do the following activities:

* Data Collection: Implementing efficient mechanisms to capture data from various sources, including user interactions, ride data, driver activities, and market trends, in real-time.
* Data Transformation and Enrichment: Applying data transformations and enrichment techniques to augment raw data and make it more informative for advanced analytics and machine learning.
* Advanced Analytics: Building tools and capabilities for data scientists and analysts to conduct advanced analytics, predictive modeling, and statistical analysis to gain deeper insights into user behavior and market trends.
* Data Visualization: Creating intuitive and interactive dashboards and visualizations to communicate data insights effectively to stakeholders at various levels within the organization.

The "Uber Data Analytics Project" represents the power of big data and advanced analytics, that aims to unlock new opportunities for innovation, enhance user experiences, and optimize business operations. The successful implementation of this project helped me not only apply my learned concepts but also gain insights from data that can help take business decisions.

**Objectives:**

* To create a data pipeline for uber data and gather insights from the data.
* Apply data transformation techniques and have the data ready in desired format.
* Use Google Cloud platform services to implement the project
* Perform analysis via Google Bigquery service and create a dashboard from our analysis.
* Imcorporating mage data tool to understand the modern data engineering tool.

**Technology and Services used:**

1. **Google Cloud -** Google Cloud is a comprehensive suite of cloud computing services offered by Google that provides organizations and individuals with a wide array of tools and resources to build, deploy, and scale applications and services on the cloud. With data centers located around the world, Google Cloud offers a reliable and powerful infrastructure to meet the ever-evolving needs of businesses, developers, and users. The platform encompasses a diverse set of services, including computing, storage, databases, machine learning, data analytics, networking, and more, all designed to empower businesses with the flexibility, scalability, and security required to succeed in today's digital landscape.
2. **GCP Cloud Storage** - Google Cloud Storage is a highly scalable and cost-effective object storage service provided by Google Cloud Platform (GCP). It offers developers and businesses a secure and reliable way to store and retrieve any amount of data in the cloud. Google Cloud Storage is designed to accommodate a wide range of use cases, from simple data backups and archiving to serving large-scale websites and applications with high availability and low latency. At its core, Google Cloud Storage is built on a distributed architecture, allowing it to store and manage vast amounts of data across multiple data centers worldwide. One of the key benefits of using Google Cloud Storage is its ease of integration with other Google Cloud services and tools. Google Cloud Storage supports a variety of data types, including structured and unstructured data, multimedia files, and large datasets. It uses a simple and straightforward RESTful API, allowing developers to interact with their data programmatically and perform operations like uploading, downloading, and managing objects with ease. Furthermore, Google Cloud Storage provides different storage classes, each designed to meet specific performance and cost requirements. These storage classes include Standard, Nearline, Coldline, and Archive, offering various levels of access times and pricing options to cater to different data access patterns and retention needs.
3. **GCP Compute Engine** - Google Cloud Compute Engine is a highly scalable and flexible Infrastructure as a Service (IaaS) offering provided by Google Cloud Platform (GCP). It enables users to deploy and manage virtual machines (VMs) on Google's vast global network, empowering businesses and developers to run their applications and workloads with ease and efficiency. Compute Engine provides a robust, reliable, and secure environment for running a wide range of applications, from simple web servers to complex, distributed computing clusters. Leveraging Google's cutting-edge technology, Compute Engine offers exceptional performance, low-latency networking, and high availability, ensuring that your applications and services remain accessible and responsive at all times.
4. **GCP BigQuery -** Google Cloud BigQuery is a fully-managed and serverless data warehouse and analytics solution offered by Google Cloud Platform (GCP). It is designed to handle large volumes of data, enabling businesses and data professionals to efficiently store, query, and analyze vast datasets in real-time. With its powerful capabilities and scalability, BigQuery is an integral part of the Google Cloud ecosystem, supporting data-driven decision-making and advanced data analytics. Key features of Google Cloud BigQuery include:

* Serverless Architecture
* Scalability
* Real-time Analytics
* Seamless Integration
* Security and Reliability
* Cost-effectiveness
* Machine Learning Integration
* Data Sharing

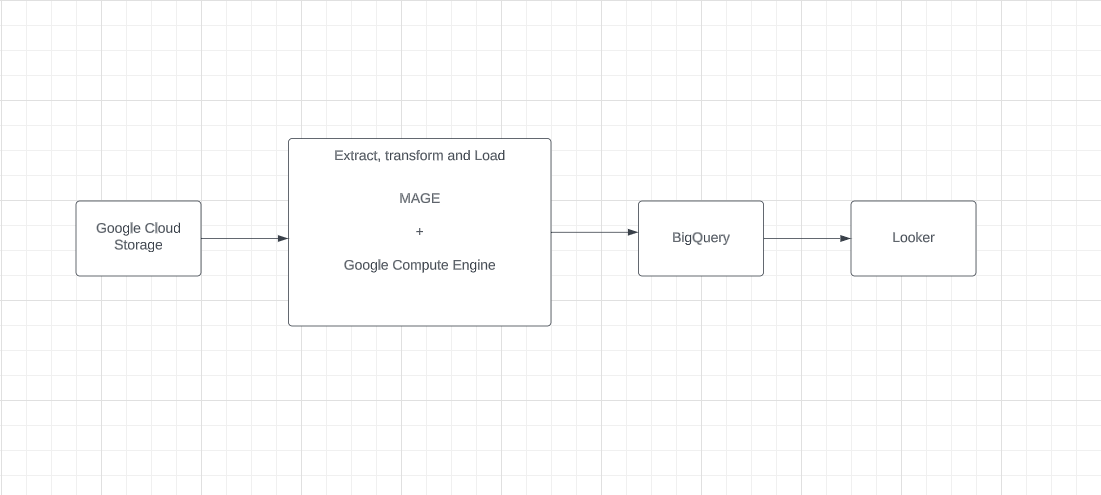
1. **Looker -** Looker is a business intelligence (BI) and data analytics platform that allows organizations to explore, analyze, and visualize their data in a user-friendly and interactive manner. It was founded in 2012 and gained popularity for its intuitive interface and powerful data exploration capabilities. Key Features of Looker:

* Data Exploration
* Data Modeling
* Interactive Visualizations
* Collaborative Environment
* Data Governance and Security
* Integrations
* Scheduled Reports
* Embedded Analytics

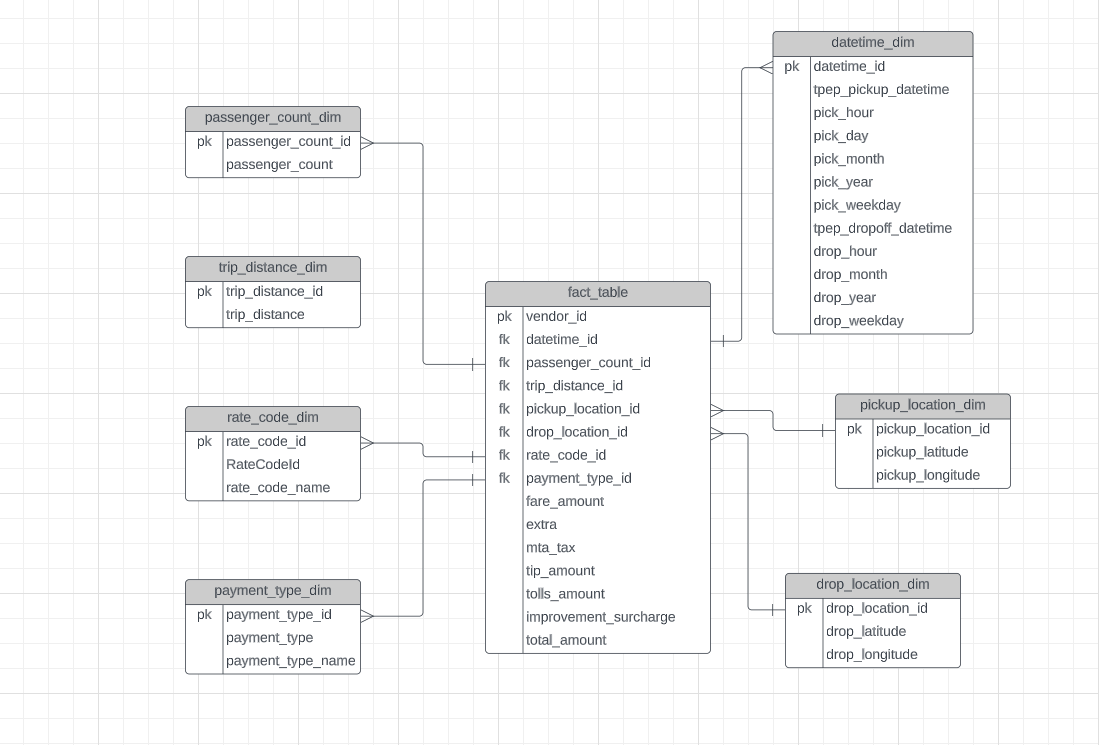
1. **MAGE -** Mage Data Tool seeks to bridge this gap by providing a comprehensive digital platform that enables mages to compile, store, and analyze vast volumes of data in a structured and accessible manner. Key features of Mage Data Tool include:

* Data Collection and Integration:
* Data Visualizatiom
* Spell Optimization
* Collaborative Research
* Security and Privacy

**Architecture:**

****

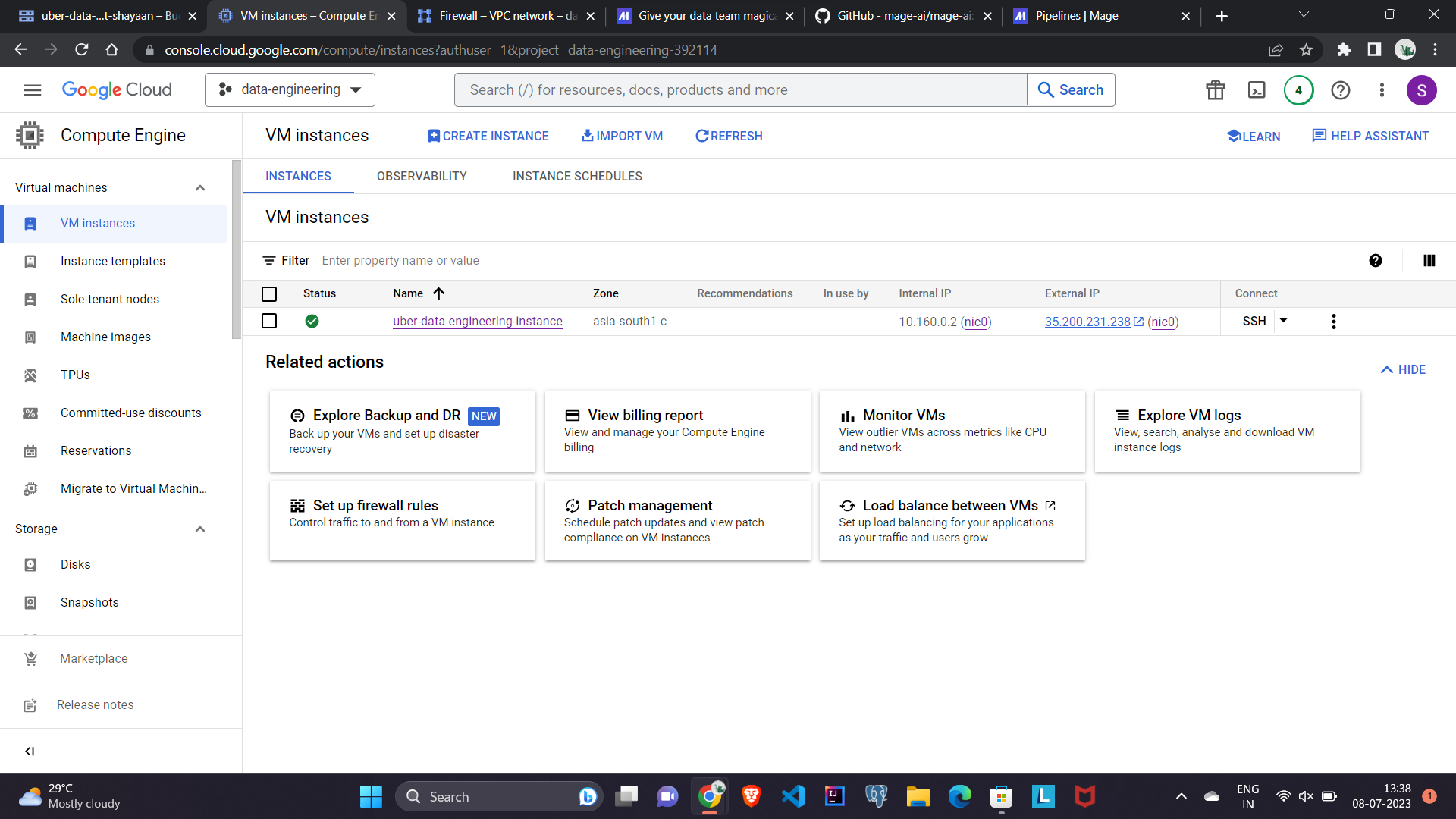
**Data Model:**

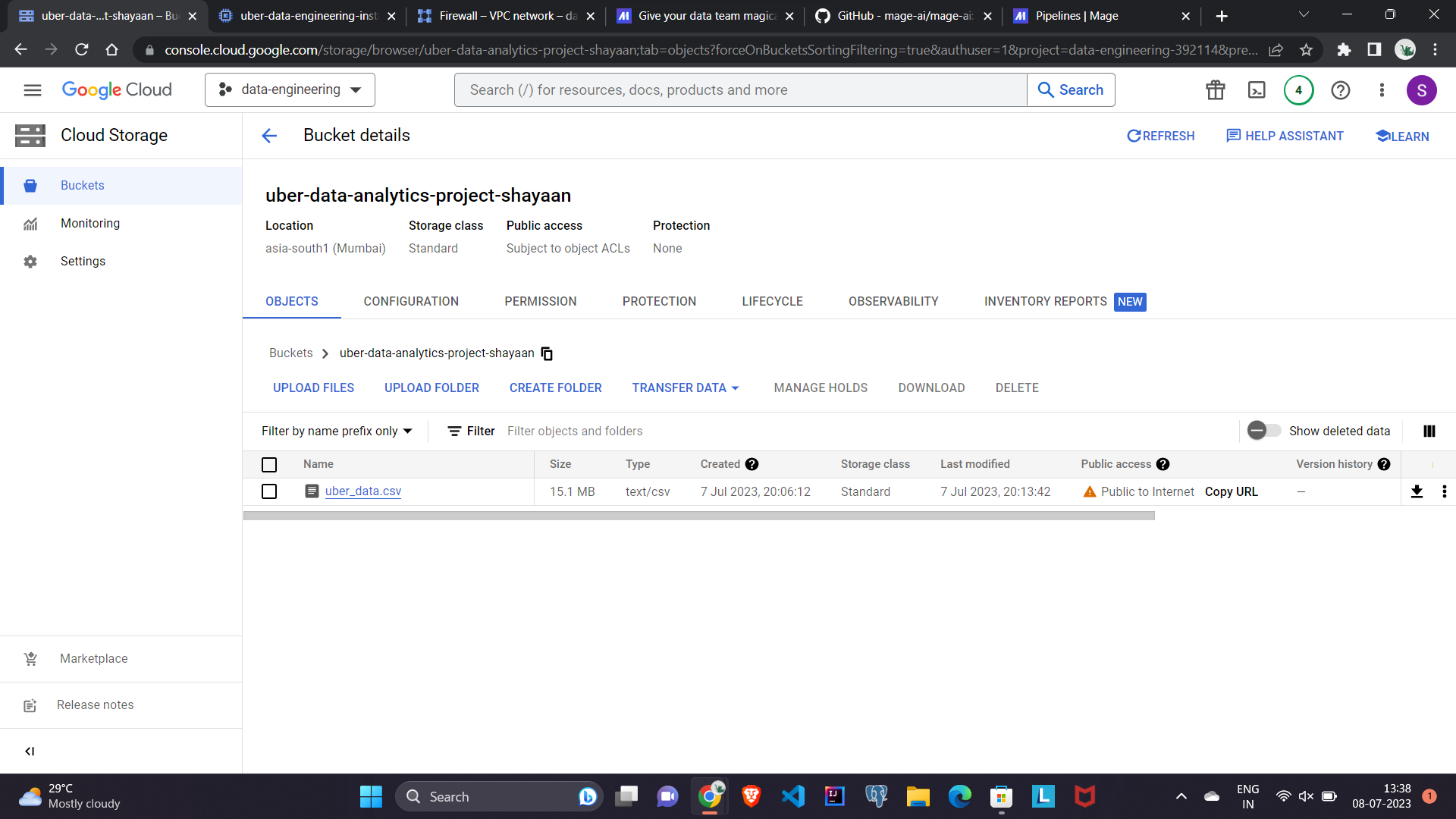
****

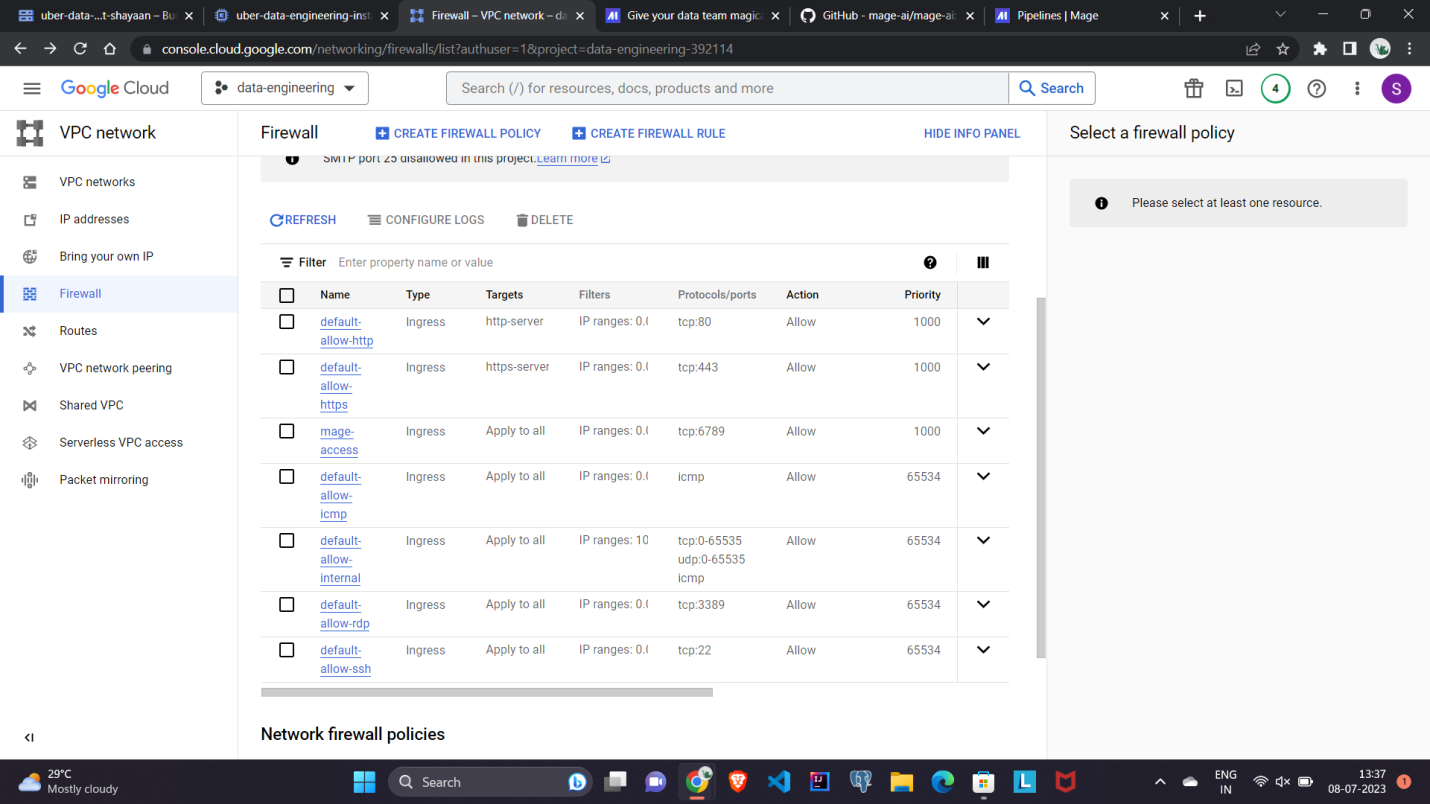
**Process:**

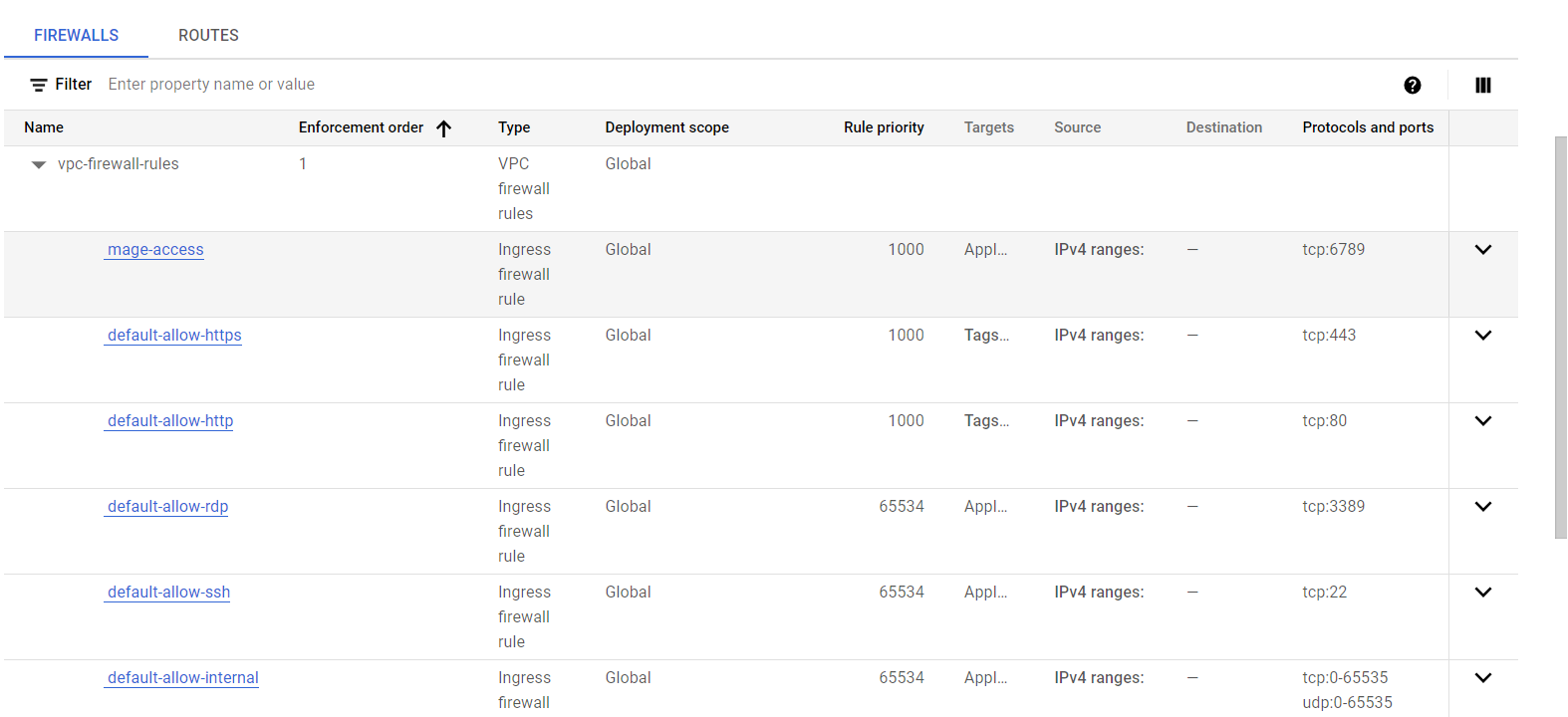
* First we will load the data into the Google Cloud Storage. We will first make a bucket and load our csv data file into the bucket. We will make the data public so we can access through the internet.
* Now we create a compute engine instance and ssh into it. In the linux terminal we connect to the mage data tool. Once the connection is successful mage data tool user interface will be available at port 6789. This will help us to create a pipeline.
* Our pipeline start with loading data from the cloud storage to a dataframe. Once done we will apply data transformation techniques using python and store the resultant data into another dataframe or we can export the data for querying.
* Once data is transformed we will export the data to Google Bigquery service and perform analysis with the help of SQL.
* Once we have gathered the required data as per our need we can create a dashboard on looker studio

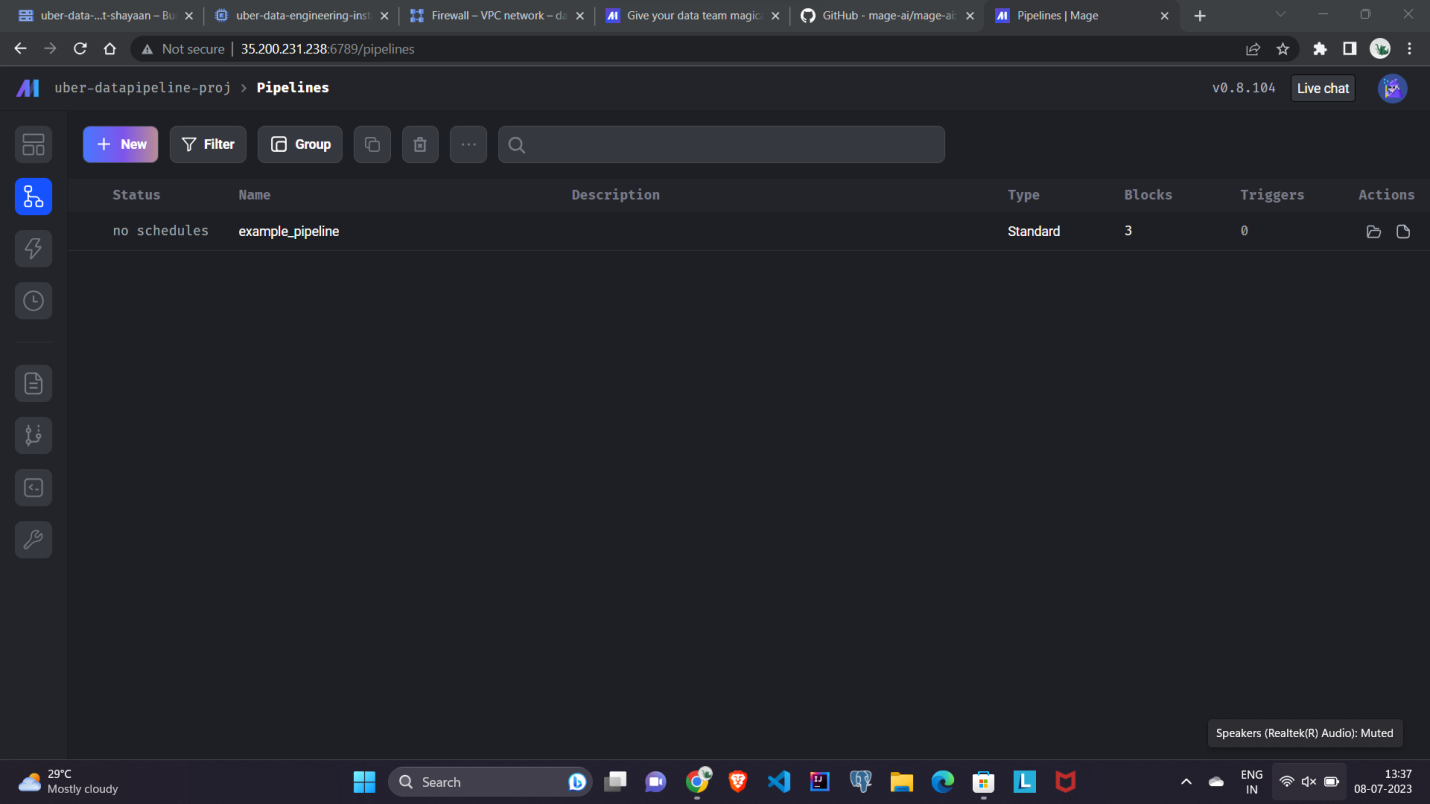
**Output:**

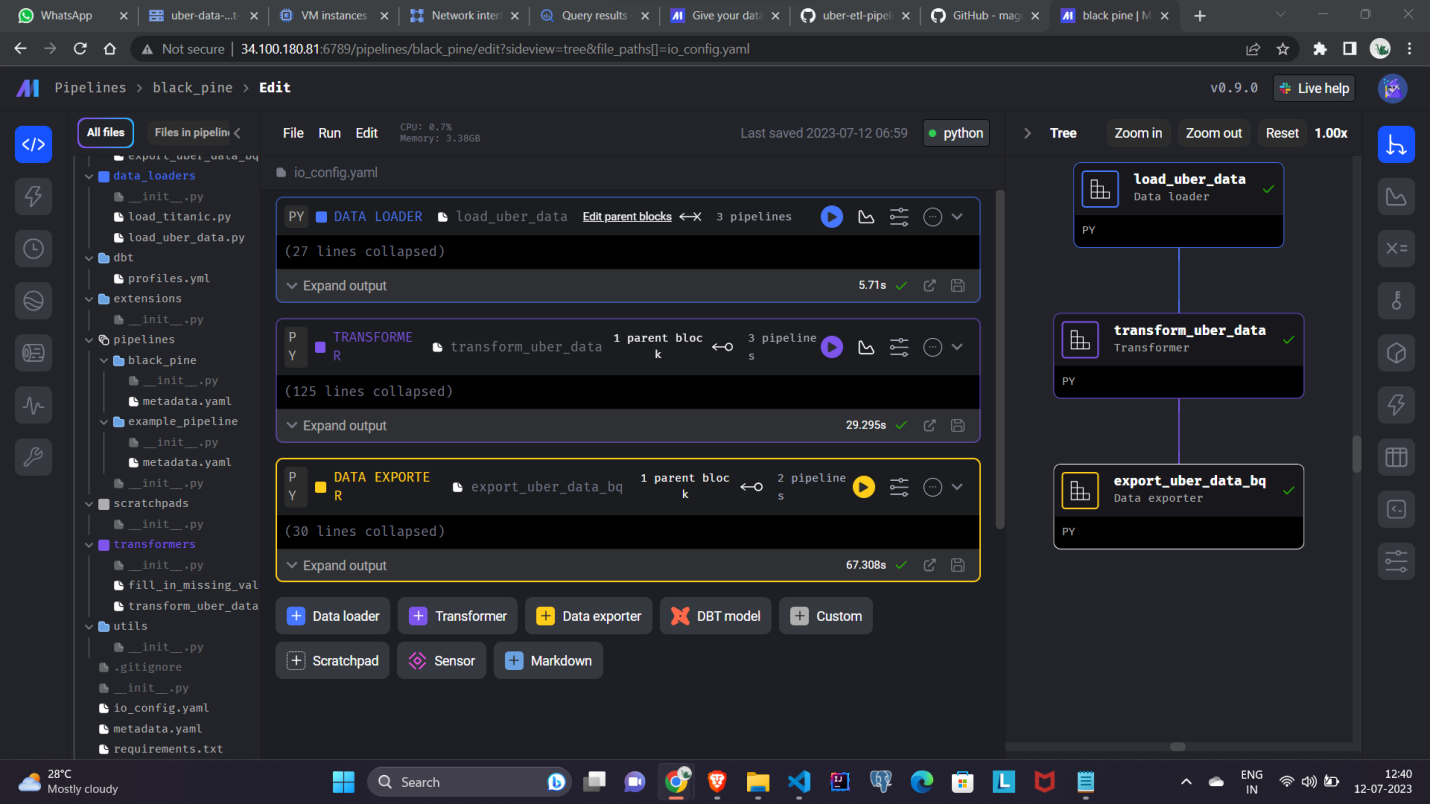
****

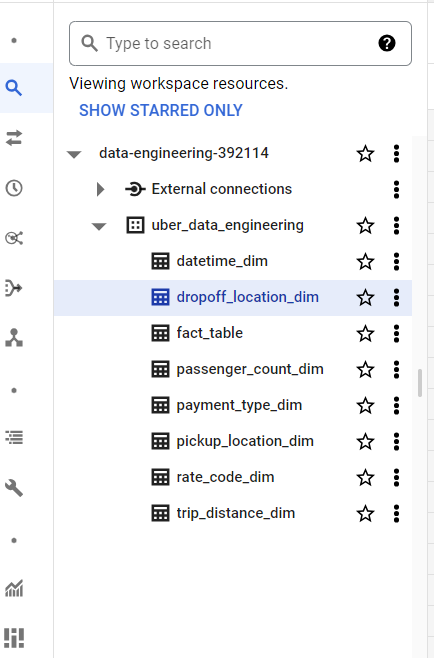
****

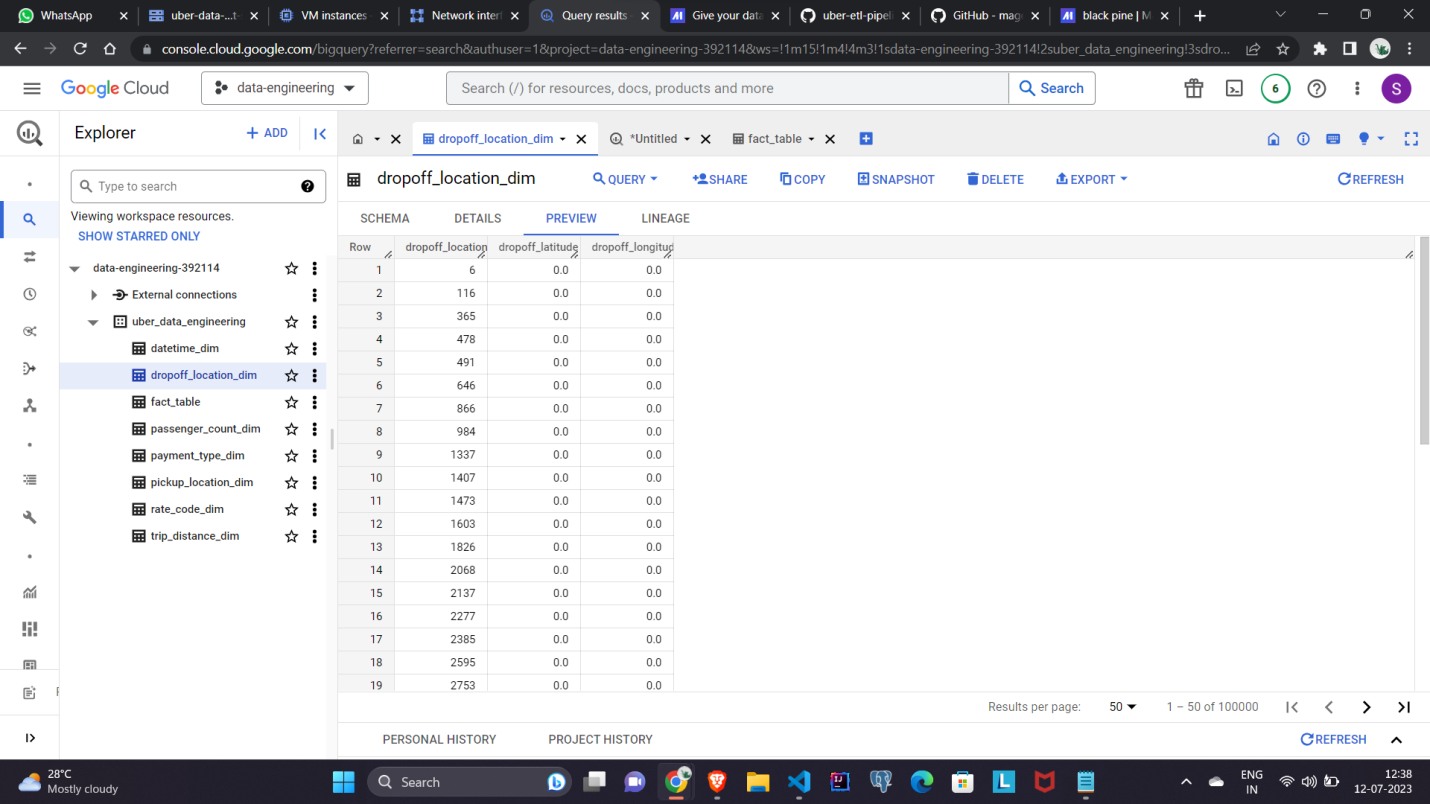
****

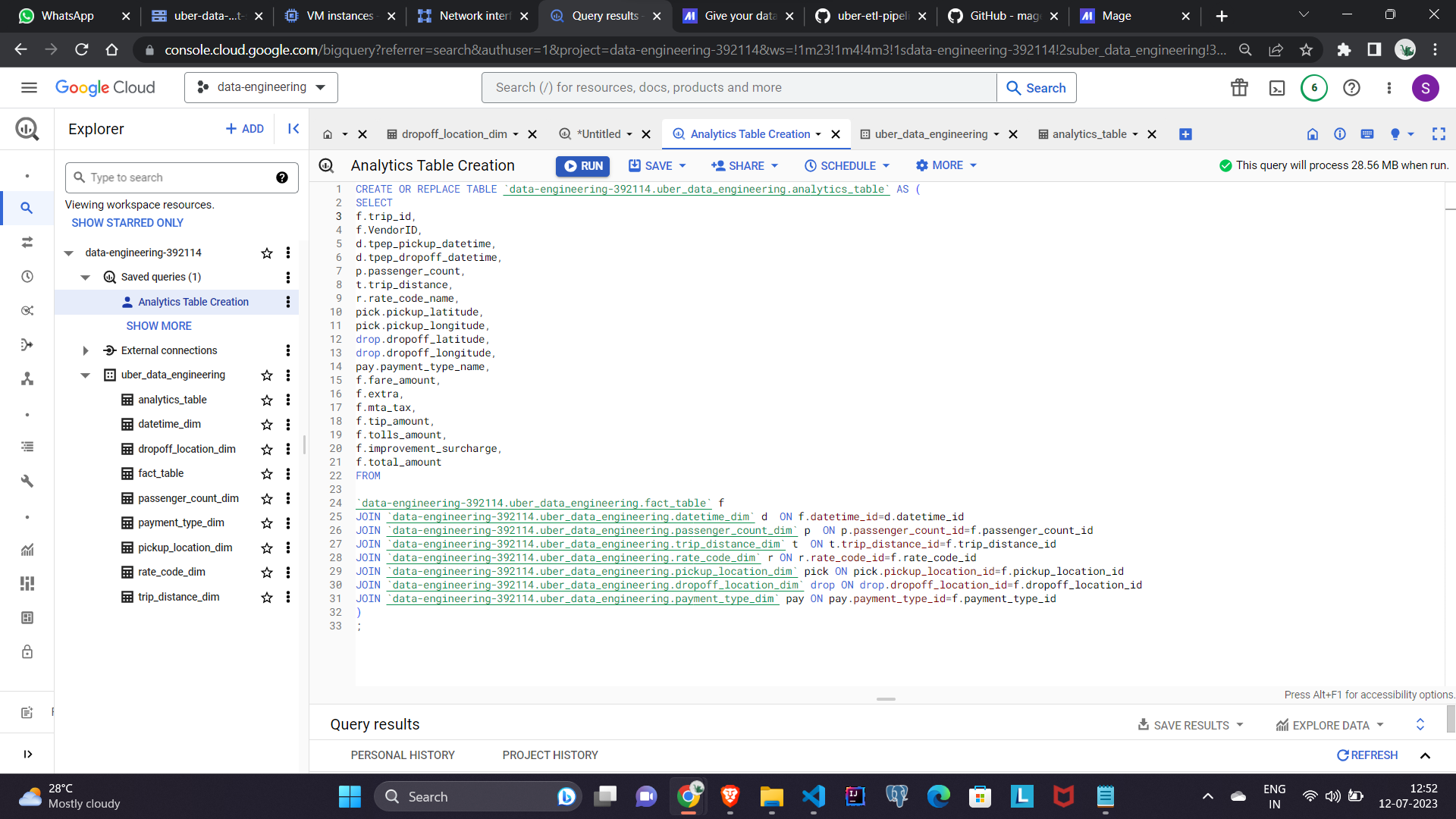
****

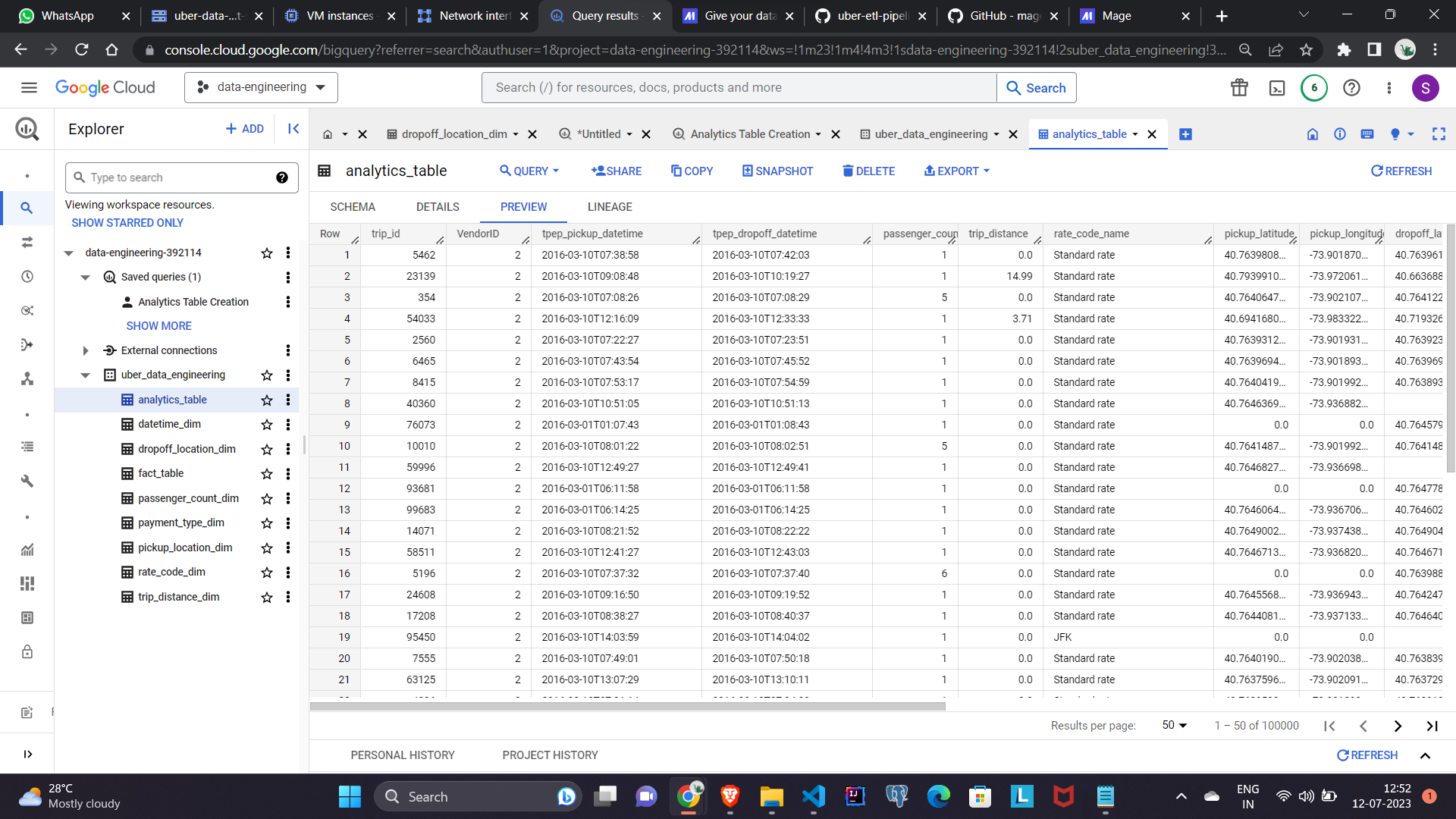
****

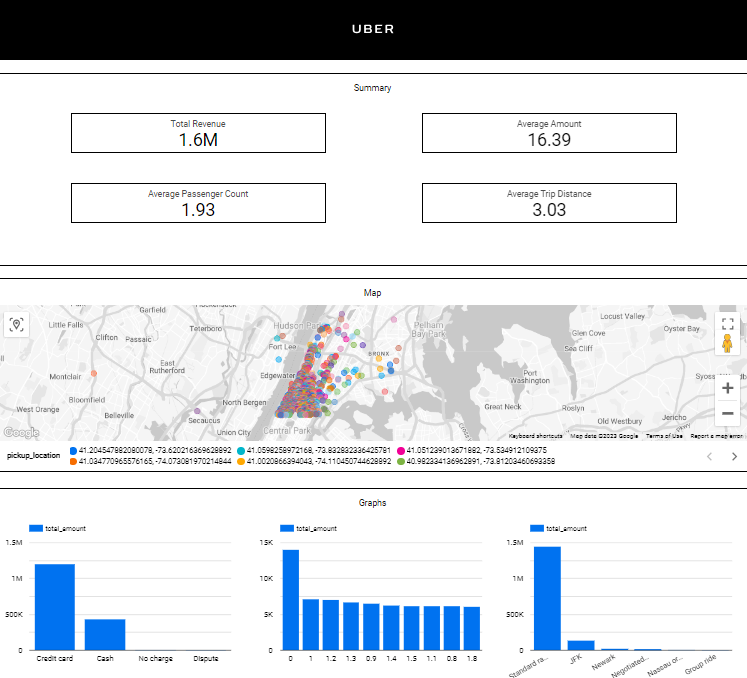
****

****

****

****

****

****