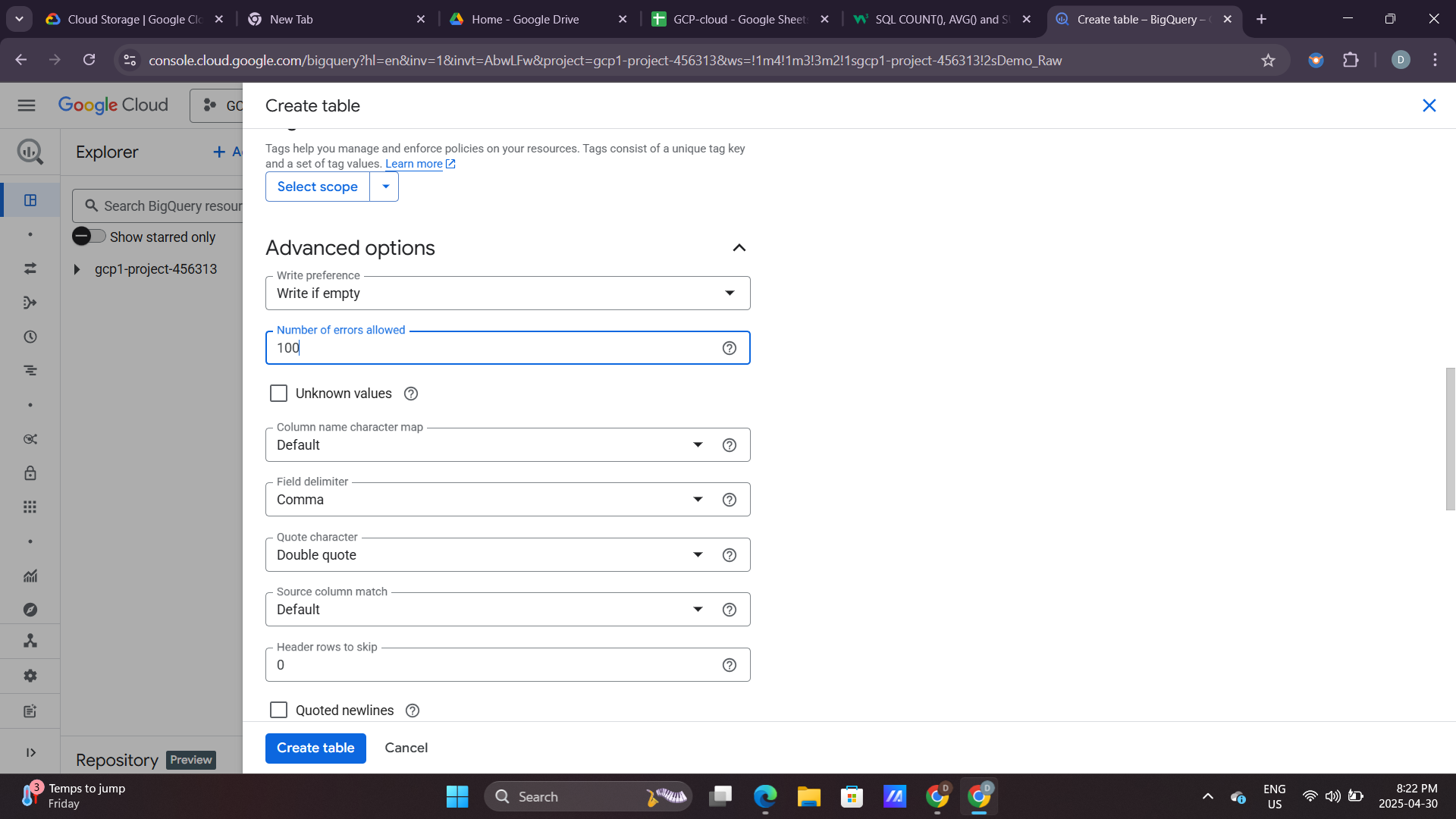
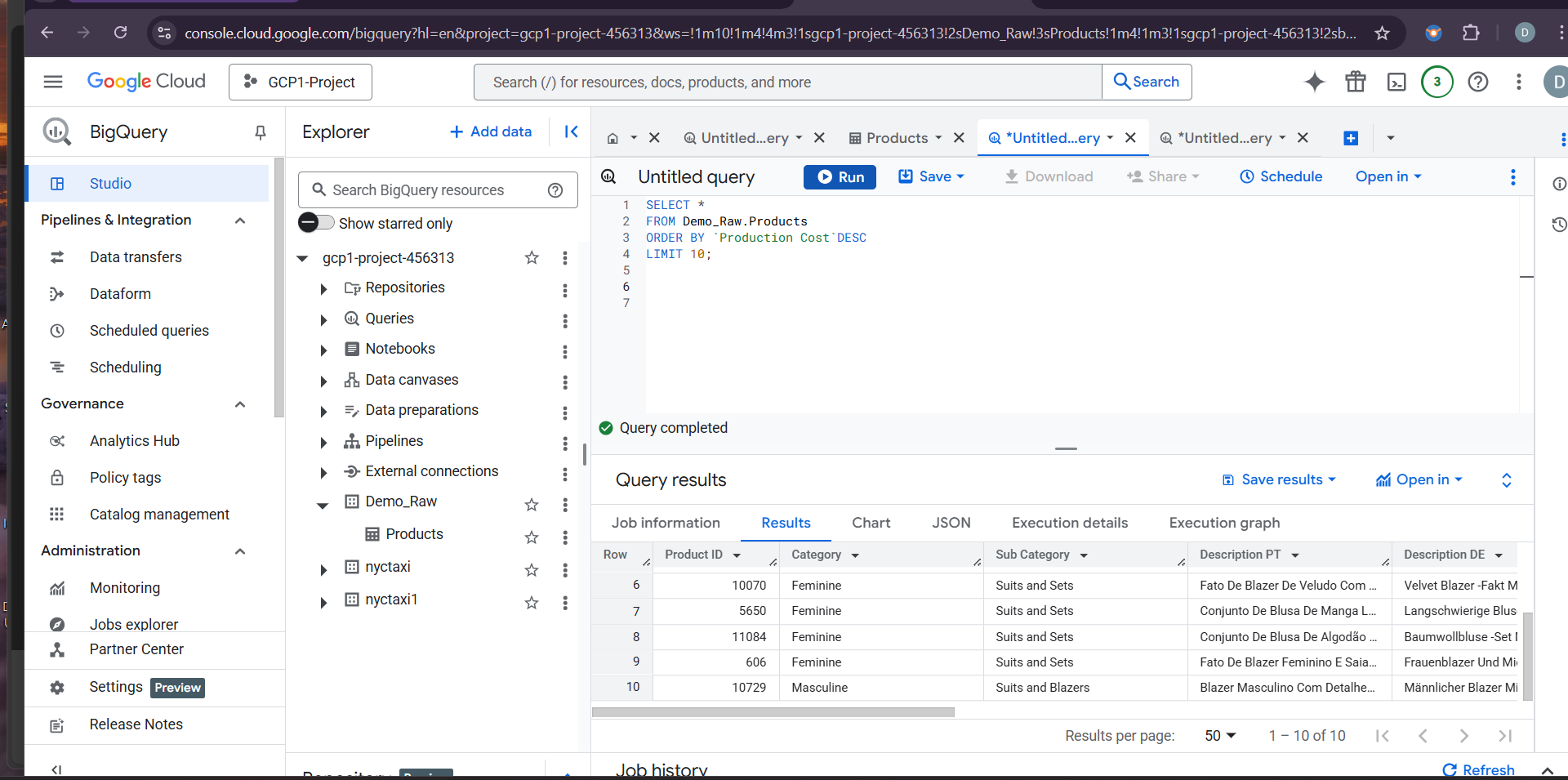
Dharma Devi 29th April HW-house usecase

Assignment 1



ORDER BY-SQL



Assignment 2: Draft Datalake-House project

Use Case: Real Estate Data Lakehouse

### **Business Objective:**

To create a centralized and scalable data platform that integrates real estate property listings, housing market trends, customer interactions, enabling real-time insights for business users, analysts, and data scientists.

### **Key Challenge Addressed:**

* Data between sales, property management, customer service, and third-party sources

**Data Sources:**

* **Property listing feeds** ( Zillow, local agencies)
* **Customer interactions** (website, app, support tickets)
* **Public housing market data** (census, interest rates, rental trends)

**Lakehouse Benefits for Housing Project:**

|  |  |  |
| --- | --- | --- |
| **Area** | **Traditional Model** | **Lakehouse Model** |
| **Data Ingestion** | Manual ETL for each system | Streaming and batch pipelines from all sources |
| **Data Storage** | Fragmented (SQL, flat files) | Centralized, schema-enforced, open format (Parquet/Delta) |
| **Analytics** | Static BI reports | Real-time dashboards, predictive analytics |
| **Governance** | Spreadsheet-driven | Metadata catalog, RBAC, audit logs |
| **ML Use Cases** | Limited, offline | Integrated model training and deploymen |

### **Use Case Examples:**

#### **1. Dynamic Property Pricing Recommendation**

* Combine historical sale prices, market trends, and property features
* Use ML models trained in the lakehouse to suggest optimal prices

#### **2. Customer 360 for Buyer Insights**

* Aggregate interactions across web, email, and in-person
* Score leads and personalize property suggestions

#### **3. Real-Time Market Intelligence Dashboard**

* Merge listing feeds with regional economic data
* Provide executives and agents live insights on hot markets

### **Technology Stack Example :**

* **Storage**: Google Cloud Storage with Delta Lake (or **BigLake** for a fully integrated lakehouse solution)
* **Compute**: Databricks on Google Cloud or **Google Cloud Dataproc** (managed Spark clusters)
* **Ingestion**: **Google Cloud Pub/Sub** (for IoT streaming), **Airbyte** (for API-based data integration), **dbt** (for transforming data within BigQuery)
* **Catalog**: **Google Cloud Data Catalog** for metadata management and governance
* **BI**: **Looker** (Google's BI tool) or **Google Data Studio** for dashboarding and reporting
* **ML**: **Vertex AI** for model training, serving, and AI-driven insights

A Data Lakehouse for a housing project provides a unified platform that integrates diverse data sources such as property listings, customer interactions and market trends. By combining the scalability of data lakes with the structured querying capabilities of data warehouses, it enables real-time analytics, predictive pricing models, and personalized customer experiences. This approach streamlines data management, enhances decision-making for real estate professionals within a single, cost-efficient architecture.

References:

<https://docs.google.com/document/d/1zTQxx2BhthUzwqU8PfEql65uMgqJoyzi0SnoI2vh0UY/edit?usp=drivesdk>

Self learning:

----------------

https://www.youtube.com/watch?v=IFMebwmDgts

https://www.youtube.com/watch?v=SrMcTinAA2M

https://www.youtube.com/watch?v=MH5M2Crn6Ag

https://www.youtube.com/watch?v=jKYkHqEwlHg