

Project Members:

Alisha Khan 2409036

Bhargavi Kadam 2409004

Simran Patel 2409006

BigQuery CloudStock Analytics

Documentation

Introduction

Google BigQuery is a fully managed, serverless data warehouse designed for efficiently storing and analyzing large datasets. This project, **BigQuery CloudStock Analytics**, demonstrates how to leverage BigQuery for stock market and real estate data analysis, SQL-based insights, and data visualization using Looker Studio. With the ability to handle massive amounts of structured and semi-structured data, BigQuery is an excellent choice for financial and real estate trend analysis.

Google Cloud & BigQuery Setup

To begin using BigQuery, users must first set up a **Google Cloud account**. Google offers a **\$300 free credit** for new users, which can be utilized to explore BigQuery services. Once the account is set up, the **BigQuery API** must be enabled through the Google Cloud Console. Next, a **Google Cloud Project** needs to be created where all datasets and tables will reside. While a billing account is required, users can take advantage of the free-tier options available.

Creating a BigQuery Data Warehouse

A **dataset** serves as the primary container for organizing financial and housing data within BigQuery. To create a dataset:

- Navigate to BigQuery in the Google Cloud Console.
- Select the project and click on "Create Dataset."
- Define a dataset name (e.g., **financial_data** or **housing_data**) and select a data location (e.g., **us-central1**).

Loading Data into BigQuery

Stock Market Data

Stock market data can be sourced from platforms such as **Yahoo Finance** or **Kaggle**. The data can be uploaded in **CSV format** using BigQuery's **Create Table** option. Users can choose to manually define schemas or let BigQuery **auto-detect** column types. Once uploaded, the data is indexed and ready for analysis.

Global Housing Market Data

For real estate analytics, we have uploaded a **global housing market dataset** that includes information on rent indices, housing prices, and economic indicators. The dataset is stored in **BigQuery under the housing_data dataset**. Data was sourced from public real estate databases and economic research institutions.

The housing dataset contains key economic and financial indicators related to real estate markets across different countries and years. It includes attributes such as the **Country** and **Year**, allowing for comparative and trend analysis. The **Affordability Ratio** measures housing affordability based on income and housing costs, while the **Construction Index** reflects real estate development activity. Economic factors such as **GDP Growth** and the **Inflation Rate** are included to understand how macroeconomic conditions impact housing trends.

The dataset also tracks **House Price Index (HPI)** and **Rent Index**, which help in monitoring property value and rental price fluctuations. Additionally, **Mortgage Rate** data provides insight into borrowing costs, influencing home affordability. Demographic factors such as **Population Growth** and **Urbanization Rate** help assess housing demand, particularly in rapidly growing urban areas. The **Record Count** represents the number of entries for a given country or year. This dataset can be used for real estate market analysis, investment decisions, urban planning, and economic policy studies by examining how various factors interact to shape housing trends worldwide.

Running Queries & Performing Analytics

BigQuery supports **SQL-based querying**, allowing users to analyze stock and housing trends. Below are some key queries for analysis:

Stock Market Queries

- **Retrieve the top 10 records from the dataset:**

```
SELECT * FROM
`your_project.financial_data.stock_table`
LIMIT 10;
```

- **Find the highest stock price for a specific company:**

```
SELECT symbol, MAX(close) AS highest_price
FROM `your_project.financial_data.stock_table`
WHERE symbol = 'AAPL'
GROUP BY symbol;
```

- **Analyze stock price trends over time:**

```
SELECT date, AVG(close) AS avg_close_price
FROM `your_project.financial_data.stock_table`
GROUP BY date
ORDER BY date DESC;
```

Real Estate Queries

- **Calculate the average rent index by country:**

```
SELECT Country,
      AVG(SAFE_CAST('Rent Index' AS FLOAT64))
FROM `housingproj.Housing.housing_data`
GROUP BY Country;
```

- **Calculate the yearly rent index by country:**

```
SELECT Year, Country,
      AVG(SAFE_CAST('Rent Index' AS FLOAT64))
FROM `housingproj.Housing.housing_data`
```

GROUP BY Year, Country;

- **Compute year-over-year growth in rent index:**

```
WITH YearlyPrices AS (  
    SELECT  
        EXTRACT(YEAR FROM SAFE_CAST('Year' AS  
DATE)) AS year,  
        Country,  
        AVG(SAFE_CAST('Rent Index' AS FLOAT64))  
AS avg_price  
    FROM `housingproj.Housing.housing_data`  
    GROUP BY year, Country  
)  
SELECT  
    Country,  
    Year,  
    avg_price,  
    LAG(avg_price) OVER (PARTITION BY Country  
ORDER BY year) AS prev_year_price,  
    ((avg_price - LAG(avg_price) OVER  
(PARTITION BY Country ORDER BY year)) /  
    LAG(avg_price) OVER (PARTITION BY Country  
ORDER BY year)) * 100 AS growth_rate  
FROM YearlyPrices;
```

Visualizing Data in Looker Studio

Looker Studio (formerly Google Data Studio) enables users to create **interactive dashboards** and visualizations using BigQuery data. To set up Looker Studio:

- Connect to BigQuery as a **data source**.
- Select the appropriate dataset and table.
- Use charts, graphs, and pivot tables to represent trends.
- Apply **filters and calculated fields** to gain deeper insights.

Based on your Looker Studio visualization, you have implemented the following:

1. Pie Chart (Market Share Distribution)

- You visualized the proportion of data across different countries using a **pie chart**.
- The **dimension used**: Country.
- The **metric used**: Record Count.
- This helps in understanding the contribution of each country to the dataset.

2. Table View (Data Summary by Country)

- You created a **table listing countries** alongside their **record counts**.
- This provides a **detailed breakdown** of the data in a structured format for quick reference.

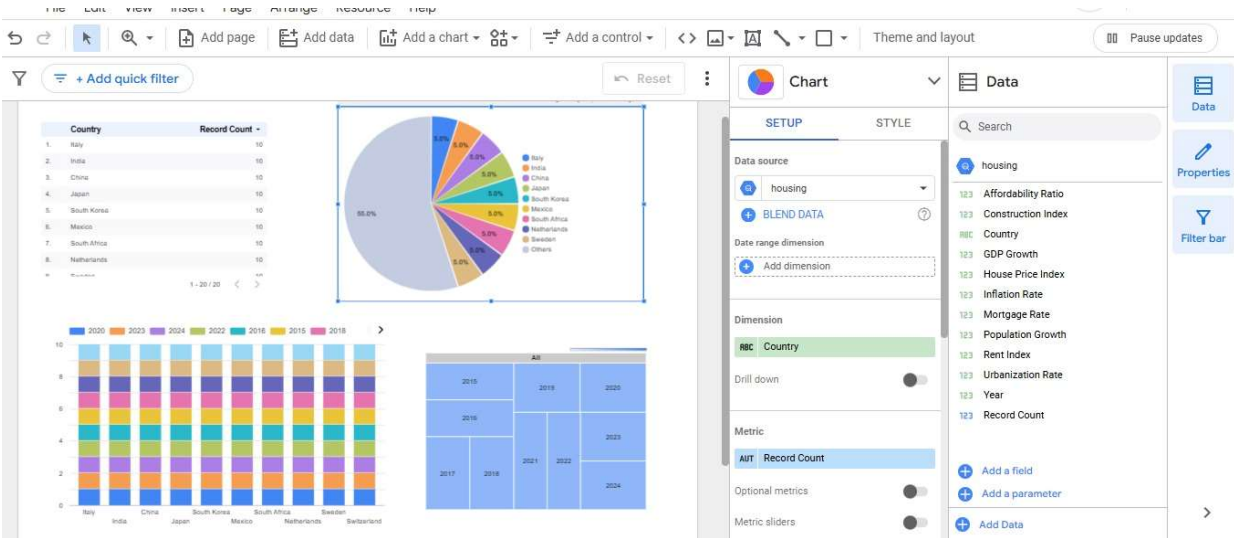
3. Stacked Bar Chart (Yearly Distribution by Country)

- A **stacked bar chart** is used to show records by country across multiple years.
- **Years are color-coded**, making it easy to compare trends over time.
- This is useful for analyzing patterns in housing or financial data.

4. Tree Map (Hierarchical Data Representation)

- You included a **tree map visualization**, which organizes data hierarchically.
- Each section represents a different year, helping users **quickly compare** distribution across time.

These visualizations together create a **comprehensive dashboard** that helps analyze the dataset from multiple perspectives.



Conclusion

BigQuery CloudStock Analytics provides a **scalable and efficient** way to analyze stock and real estate market data. By leveraging **BigQuery's SQL engine, Looker Studio visualizations, and automated data ingestion**, users can generate powerful insights into financial and housing trends. Future enhancements could include **machine learning models for price predictions** and **automated alerts for stock movements**.