Hackathon 1 - GDP Trend and Growth Spike Detector

Name : K Guru Charan

REG NO : RA2211026010141
DEPT : CSE AI &ML

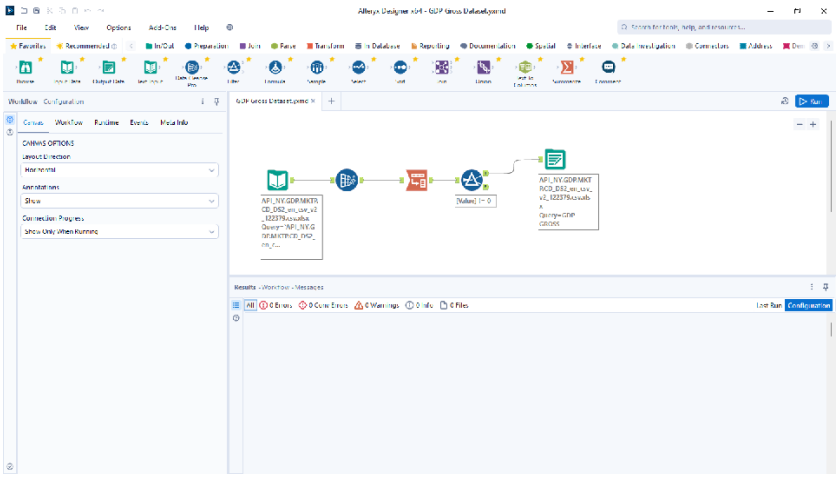
1.Preprocessing the data using alteryx

🔧 **Alteryx Workflow:** GDP Gross Dataset Preparation

- **Workflow Name:** GDP Gross Dataset.yxmd
- Purpose: To ingest, clean, and preprocess global GDP data for trend and anomaly analysis (YoY spikes/drops).
- **Data Sources:**
  - Connected to Excel/CSV file: API_NY.GDP.MKTP.CD_DS2_en_csv_v2_122379.csv.xlsx
  - Query: GDP Indicator – "API_NY.GDP.MKTP.CD_DS2"
- **Key Tools Used:**
  - 🟢 Input Data Tool: Reads raw GDP data from external Excel source.
  - 🔵 Select Tool: Keeps only required fields (Country Code, Year, GDP Value).
  - 🟡 Formula Tool: Performs calculated transformations (e.g., renaming, typecasting).
  - 🔴 Filter Tool: Filters out zero or invalid GDP values using [Value] != 0.
  - 🔷 Output Data Tool: Writes cleaned data to an output file for further use in SQL/Power BI.
- Cleaning Logic:
  - Removed rows with GDP value = 0 to ensure accurate YoY calculations.
  - Standardized column types (e.g., year as integer, GDP as float).
  - Flattened file structure to align with normalized SQL schema later.
- Output Dataset:
  - Cleaned, structured GDP table containing: Country_Code, Year, and GDP (USD).
- Integration:

- Prepares data for ingestion into the gross_gdp table in the SQL database.
- Serves as the foundation for subsequent YoY growth analysis and anomaly flagging.

- Run Status: ✅ No errors or warnings reported.



2.a Schema

Database Schema for GDP Trend Analysis

1. Table: gross_gdp

Country_Code (PK, TEXT)

Country_Name (TEXT)

Indicator_Name (TEXT)

Indicator_Code (TEXT)

Year (INTEGER)

GDP (REAL)

2. Table: gross_gdp_yoy

ID (PK, INTEGER AUTOINCREMENT)

Country_Code (FK → gross_gdp.Country_Code, TEXT)

Year (INTEGER)

GDP (REAL)

Previous_Year_GDP (REAL)

YoY_Growth (REAL)

Growth_Flag (TEXT)

(Values: 'Spike', 'Drop', 'Stable', 'No Data')

3. Table: gdp_country_summary

Country_Code (PK, FK → gross_gdp.Country_Code, TEXT)

Average_YoY_Growth (REAL)

Total_Spikes (INTEGER)

Total_Drops (INTEGER)

4. Table: gdp_yearly_summary

Country_Code (FK → gross_gdp.Country_Code, TEXT)

Year (INTEGER)

Total_Records (INTEGER)

Total_Spikes (INTEGER)

Total_Drops (INTEGER)

Total_Neutral (INTEGER)

Sum_YoY_Growth (REAL)

Avg_YoY_Growth (REAL)

Relationships :

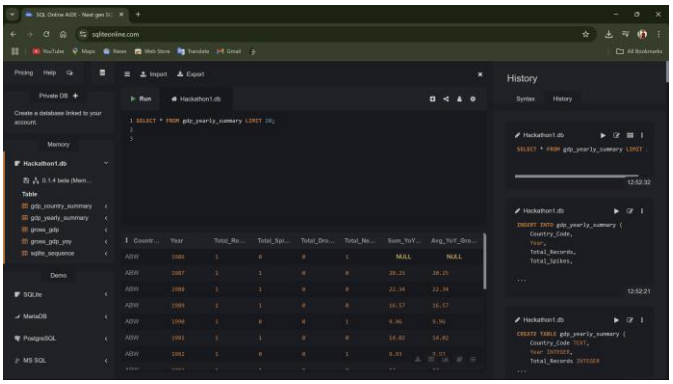
gross_gdp_yoy.Country_Code → gross_gdp.Country_Code

gdp_country_summary.Country_Code → gross_gdp.Country_Code

gdp_yearly_summary.Country_Code → gross_gdp.Country_Code

2.b Schema Diagram

3.SQL Queries

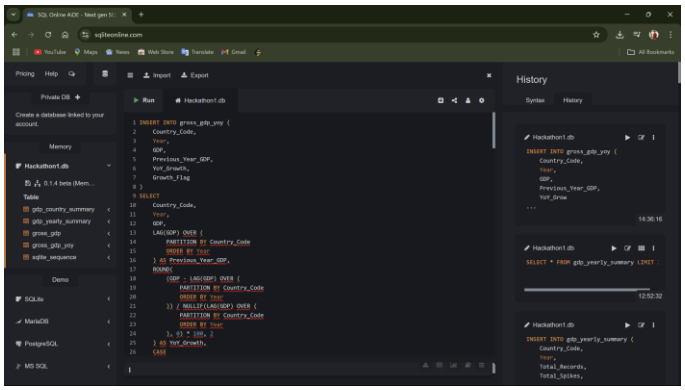


➔ Create the new table to hold YoY Growth & Flags

```
CREATE TABLE gross_gdp_yoy (
    ID INTEGER PRIMARY KEY AUTOINCREMENT,
    Country_Code TEXT,
    Year INTEGER,
    GDP REAL,
    Previous_Year_GDP REAL,
    YoY_Growth REAL,
```

```
    Growth_Flag TEXT,
    FOREIGN KEY (Country_Code) REFERENCES gross_gdp(Country_Code)
);
```

➔ Insert computed YoY Growth and flags



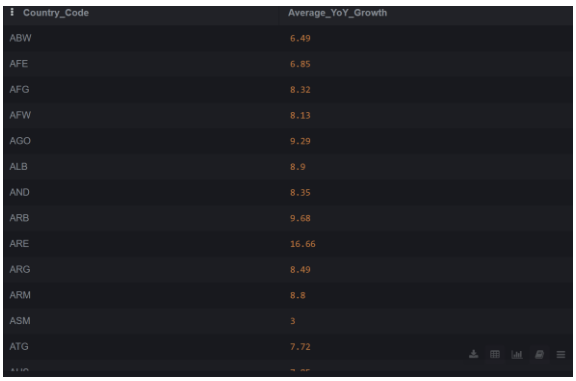
Output :

ID	Country_C...	Year	GDP	Previous_Y...	YoY_Growth	Growth_Flag
2	ABW	1987	487709497.2	405586592.2	20.25	Spike
3	ABW	1988	596648044.7	487709497.2	22.34	Spike
4	ABW	1989	695530726.3	596648044.7	16.57	Spike
5	ABW	1990	764804469.3	695530726.3	9.96	Normal
6	ABW	1991	872067039.1	764804469.3	14.02	Spike
7	ABW	1992	958659217.9	872067039.1	9.93	Normal
8	ABW	1993	1083240223	958659217.9	13	Spike
9	ABW	1994	1245810056	1083240223	15.01	Spike
10	ABW	1995	1320670391	1245810056	6.01	Normal
11	ABW	1996	1379888268	1320670391	4.48	Normal
12	ABW	1997	1531843575	1379888268	11.01	Spike
13	ABW	1998	1665363128	1531843575	8.72	Normal
14	ABW	1999	1722905028	1665363128	3.46	Normal

➔ Average YoY Growth per country:

```
SELECT
    Country_Code,
    ROUND(AVG(YoY_Growth), 2) AS Average_YoY_Growth
FROM
    gross_gdp_yoy
WHERE
    YoY_Growth IS NOT NULL
GROUP BY
    Country_Code;
```

ouput :

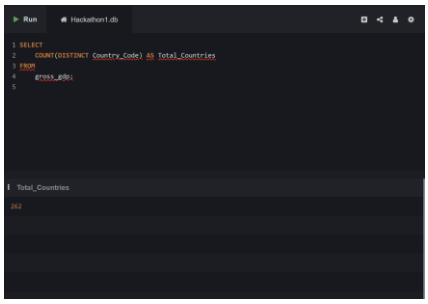


Country_Code	Average_YoY_Growth
ABW	6.49
AFE	6.85
AFG	8.32
AFW	8.13
AGO	9.29
ALB	8.9
AND	8.35
ARB	9.68
ARE	16.66
ARG	8.49
ARM	8.8
ASM	3
ATG	7.72

➔ Total Countries:

```
SELECT
    COUNT(DISTINCT Country_Code) AS Total_Countries
FROM
    gross_gdp;
```

Output :

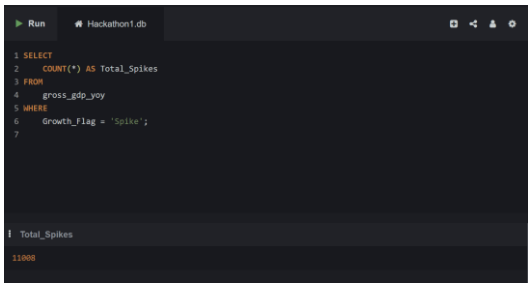


➔ Total Drops:

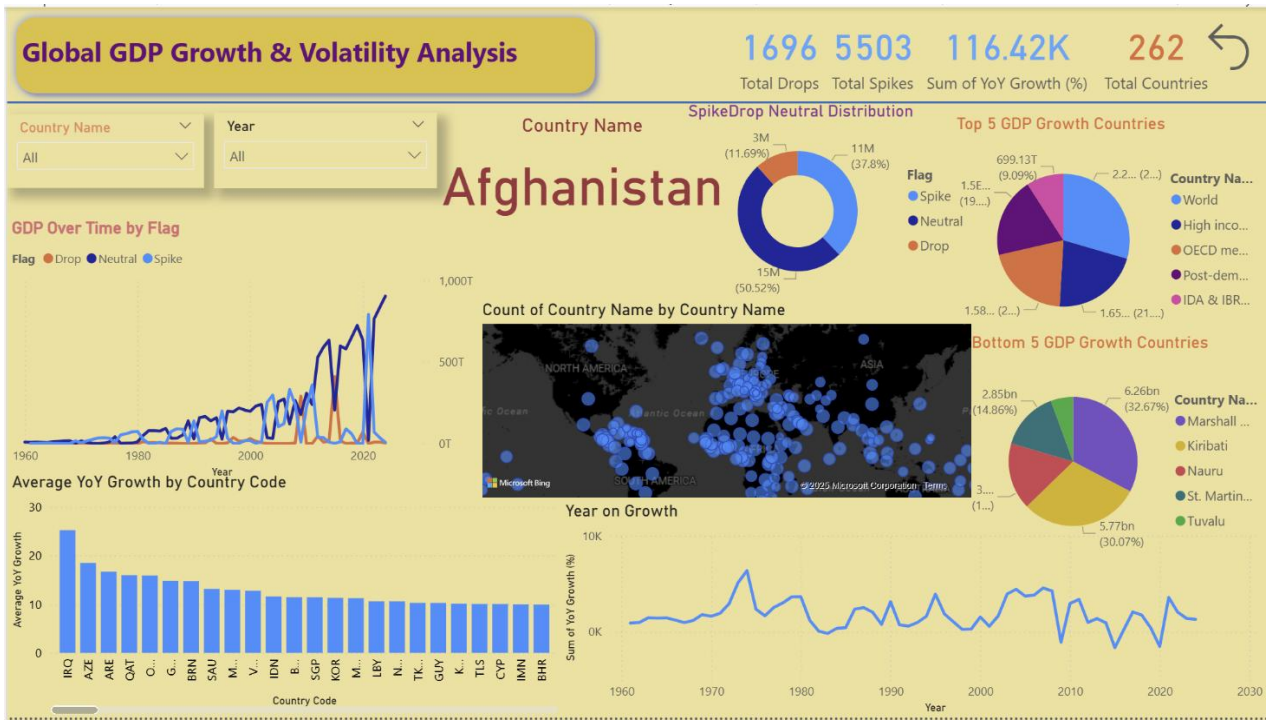
```
SELECT
    COUNT(*) AS Total_Drops
FROM
    gross_gdp_yoy
WHERE
    Growth_Flag = 'Drop';
```

Output :

➔ Total spikes :



4.Power BI Dashboard



GDP INSIGHTS :

1. Global Volatility Indicators

- **Total Spikes** (5503) far exceed Total **Drops (1696).**Global GDP trends show more positive growth years than negative; this suggests general economic expansion.
- **116.42K%** is the sum of Year-on-Year (YoY) Growth. This indicates cumulative global GDP growth over recorded years..

2. Spike/Drop/Neutral Distribution

- **Spikes (50.5%)** dominate. A majority of the data years show GDP increases of over **10%**. This indicates frequent economic booms in some nations.
- **Drops (11.7%)** are minimal. Recessions or sharp declines are less common worldwide. .

3. Top vs Bottom GDP Growth Countries

- **Top 5 Growth Contributors**:
    - **Includes macro-regions** such as World, OECD members, and IDA and IBRD groups.
    - This means the analysis looks at both countries and economic blocs to identify overall trends.
- **Bottom 5 Growth Countries**:
    - - Include **Marshall Islands, Kiribati, Nauru, Tuvalu, and St. Martin.**
    - - These are mostly small island nations. They are vulnerable to changes because of their limited economic bases.

4. Trends Over Time (1960–2023)

- **"GDP Over Time by Flag"** shows blue spikes that concentrate after 2000. There is a noticeable increase in economic volatility in recent decades, especially after 2010.
- The periods following **2008 and 2020** show sharp drops. These are likely due to the Global Financial Crisis and **the COVID-19** pandemic, respectively.

5. Country-specific Growth Patterns

- **Bar chart of "Average YoY Growth by Country Code"** shows:
    - **IRAQ (IRQ) leads** with nearly **30% average** YoY growth, possibly due to periods of rebuilding after the war.
    - Other countries with high growth include **Azerbaijan (AZE) and Qatar (QAT);** both are rich in oil and developing quickly.

6. Geo Distribution

- **Global map visualization** confirms widespread data coverage across all continents, supporting the "global" nature of the study.

➔ PowerBI Wireframes

Top Section :

Element	Type	Placement
Country Name	Dropdown Slicer	Left
Year	Dropdown Slicer	Next to Country
Total Drops	KPI Card	Top right
Total Spikes	KPI Card	Top right
Sum of YoY Growth %	KPI Card	Top right
Total Countries	KPI Card	Top right

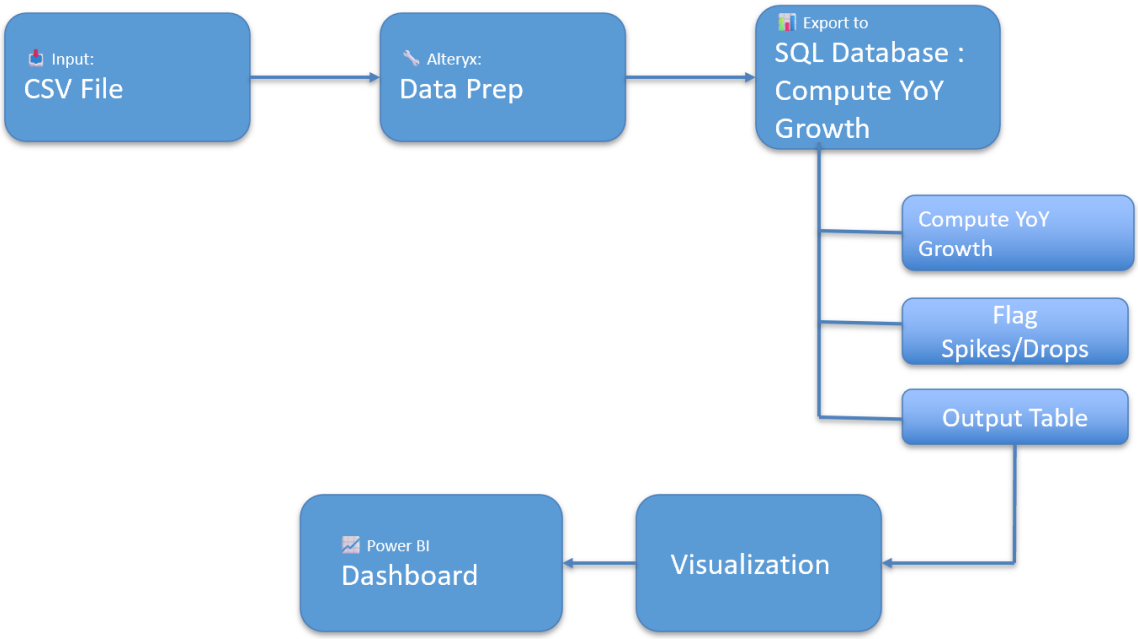
Middle Section : key visual

Visual Type	Title	Notes
Line Chart	GDP Over Time by Flag	Year on X-axis, YoY % on Y-axis, colored by Flag (Spike, Drop, Neutral)
Donut Chart	Spike/Drop/Neutral Distribution	Distribution of flag types
Pie Chart	Top 5 GDP Growth Countries	Based on cumulative YoY growth
Pie Chart	Bottom 5 GDP Growth Countries	Based on lowest YoY growth
Map (Filled/Bubble)	Count of Country by Location	Bubble size or color = # of years with data or spike/drop counts

Bottom Section : deeper trends

Visual Type	Title	Description
Bar Chart	Average YoY Growth by Country Code	Horizontal or vertical layout, sorted by growth
Line Chart	YoY Growth Trend Over Years	Total sum of YoY % globally by year

5. DATA PIPELINE DIAGRAM



Github Link : https://github.com/guruc267/GDP-Trend-and-Growth-Spike-Detector/tree/main