**RevInflation**

**(Data Engineering Project)**

Center of Excellence

27-Oct-2023

Version 1.0

Copyright ©: 2023, Revature.

​​

Table of Contents

[Application Overview 3](#_Toc2142547784)

[Core Functional Requirements 3](#_Toc1525735439)

[Standard Functional Scope 3](#_Toc503760079)

[Definition of Done 4](#_Toc735631391)

[Competency wise scoping 4](#_Toc813891622)

[Non-Functional Expectations 6](#_Toc1392265130)

[Source Data Location 6](#_Toc917099209)

​

​

​

​

​​

​​

​​

​

​

​

​

​

​​

​

​

​

​

​

​​

# Application Overview

This RevInflation aggregates major inflation metrics across countries and years into a unified dataset. It provides interactive visualizations to analyze inflation rates, trends and economic patterns. Users can filter by country, development status, and time period. Advanced analytics identify correlations, clusters, forecasting models, and anomaly detection for inflation data. The goal is to uncover insights that explain inflation dynamics across global economies over time.

# Core Functional Requirements

As a User, I want to:

**Data Aggregation**

1. Integrate the Energy Consumer Price Inflation, Food Consumer Price Inflation, Headline Consumer Price Inflation, Official Core Consumer Price Inflation, and Producer Price Inflation for different countries.

**Data Visualization**

1. Visualize a dashboard that provides information about the Energy Consumer Price, Food Consumer Price, Headline Consumer Price, Official Core Consumer Price, and Producer Price Inflation over the years for various countries. Provide an option to filter based on the country and the type of Inflation. Rank the countries based on the Inflation rate.

**Data Analysis**

1. Identify the countries with the highest and lowest inflation rates.
2. Identify the periods where the inflation is high.
3. Identify the period where the inflation is low.

**Data Mining**

1. **Identify the recession and inflation patterns over the years for various countries. This data can be used to identify the causes that might have led to the recession and Inflation.**
2. **Compare the inflation rate between the developed, developing counties, and least developing countries.**

# Standard Functional Scope

1. Ingesting and processing various financial data feeds.
2. Transforming data into analytical models.
3. Providing interactive dashboards, queries, and reports for analysis.
4. Implementing administrative controls for user management and data governance.

This scope ensures efficient data handling, advanced analysis, and user-friendly tools, facilitating seamless transformation of raw financial data into valuable insights for informed decision-making.

# Definition of Done

1. Working application demonstration.
2. Sharing the associates’ code repo for technical review with:

* Architecture
* data models
* ETL documentation

# Competency wise scoping

|  |  |  |
| --- | --- | --- |
| **Competency** | **Project Type** | **Expectations** |
| Python, SQL  Scala, SQL | REST API service | **Framework Specific**   1. Ensure the appropriate APIs are used for any of the API calls. 2. Ensure the routing is centrally configured 3. Best practices & design patterns are to be followed.     **Validation and Error Handling:**   1. Validate the inputs for their types and format. 2. Display functional-related user messages (either for input/error/output) - no system error codes or SQL error codes. 3. Handle the exceptions and errors gracefully.     **Logging:**   1. Ensure the application is using proper logging framework and methods. 2. Ensure the application’s log level is configured using configuration files so that it can be changed without changing the code. 3. Also ensure that the application logging is configured to output to the mentioned log file.     **Testing**:   1. Ensure sufficient test cases are written using appropriate testing frameworks. 2. Ensure the code coverage closed to be 80%     **Security**:   1. Ensure the SQL injection threat is taken care. 2. Ensure the CORS restriction is applied, if applicable. 3. Ensure that the secrets are stored as environment variables using secure credential storage.     **Coding Standard:**   1. Use the industry coding standards and conventions. 2. Modular based code development for better reusability. 3. Ensure proper usage of resource objects such as database connectivity objects to avoid resource leakages. 4. Ensure proper usage of design patterns and application layering (such as Business Service, DAO Layer etc.) wherever applicable. |
| Spark  Hadoop, Hive | Data Science | **Data Preprocessing:**     1. Load the inflation dataset from a reliable source. 2. Handle missing data, outliers, and any data quality issues appropriately. 3. Perform data cleansing, normalization, and transformation to prepare it for analysis.     **Exploratory Data Analysis (EDA):**     1. Generate descriptive statistics to understand the dataset's basic characteristics. 2. Visualize the relevant attributes in the data set. 3. Identify patterns, correlations, and anomalies in the data through various charts and graphs.   **Data Visualization:**     1. Design interactive visualizations (line charts, bar charts, pie charts, etc.) to represent coal production trends, energy source distribution, and other relevant insights.      1. Ensure the visualizations are user-friendly, allowing users to explore and interact with the data to gain a deeper understanding.     **Reporting and Presentation:**     1. Generate detailed reports summarizing the findings from the data analysis and mining processes.      1. Create presentations with clear explanations of insights, trends, and correlations for stakeholders' understanding.     **User Interaction and Exploration:**     1. Provide user interfaces or dashboards that allow stakeholders to interact with visualizations and customize views based on their interests.      1. Enable filtering, sorting, and comparison functionalities within the visualizations to facilitate data exploration. |
| Azure  AWS  GCP | ETL Pipeline | **Deployment artifacts:**   1. The deployment artifacts should be minified and obfuscated if required.   **Pipeline Orchestration**     1. Scheduled ETL jobs monitor, orchestrate and restart flows 2. Job monitoring and alerts for data quality and SLAs 3. Scalable compute resources based on workload   **Scalability and Performance:**     1. Design the ETL pipeline to handle large volumes of stock data efficiently. 2. Optimize data processing and transformation to minimize execution time.   **Monitoring and Alerts:**     1. Set up monitoring for ETL job status, performance metrics, and data quality. 2. Configure alerts to notify stakeholders about failures or anomalies in the ETL process. |

# Non-Functional Expectations

* Application development should use version control systems (e.g., Git) to manage the project codebase and facilitate collaboration.
* Application development is supposed to follow the Scrum process.

# Source Data Location

The data can be acquired from Kaggle using:

* Link to the csv file [Global Dataset of Inflation.csv](https://revature0.sharepoint.com/:x:/r/sites/trainers/Shared%20Documents/Center%20of%20Excellence/Project%20Operations/Data%20Engineering/Data%20Source/Global%20Dataset%20of%20Inflation.csv?d=w94ed7fe8869844968ab50647478ddfea&csf=1&web=1&e=kNFuG1)