Assignment 2 - Data Ingestion Pipeline for Time Series data

# Purpose:

Through this assignment, we intend to design data ingestion and model pipeline for Financial time series data using FRED API.

# Dataset:

Data for different time series data models needs to be ingested and used further for modeling purposes. Data has been taken from FRED website.

Rates: 1Mo CMT, 3Mo CMT, 1yr CMT, 2yr CMT, 5yr CMT, 10yr CMT, 30yr CMT

Inflation: CPI (MoM), CPI (YoY), CPE (MoM), CPE (YoY)

Economy: GDP (nominal), GDP (real), Civilian Unemployment Rate

Credit Spreads: ICE BofAML US Corporate BBB Effective Yield – 10yr CMT, ICE BofAML US Corporate BBB Effective Yield, ICE BofAML US Corporate AA Effective Yield, ICE BofAML US Corporate A Effective Yield

Agricultural Real Estate: Delinquencies on Farmland Loans (DRFLACBN), Charge-Off on Farmland Loans (CORFLACBN)

# 

# Processes Involved

## Data Ingestion:

We developed a python script wherein we fetched data for different time series models , store in a CSV and ingested data into S3 bucket - info7374-ejfc-corp. Separate folders for different models are inserted into S3 bucket.

## Data Processing and Staging:

As a part of data processing , we had to convert raw data into JSON format and store in Processed folder of S3 bucket.

Developed DAG script and scheduled fetching the data and only deltas were fetched and not the complete data.

### Github Link : <https://github.com/akashkatakam/datapipelining_twitter/tree/master/ejcorp_assignment_2>

# References:

* <https://fred.stlouisfed.org/>
* <https://news.research.stlouisfed.org/2019/10/teaching-monetary-policy-bring-fred-into-the-classroom-october-2019/#Using-GeoFRED-to-Compare-Federal-Reserve-Districts>