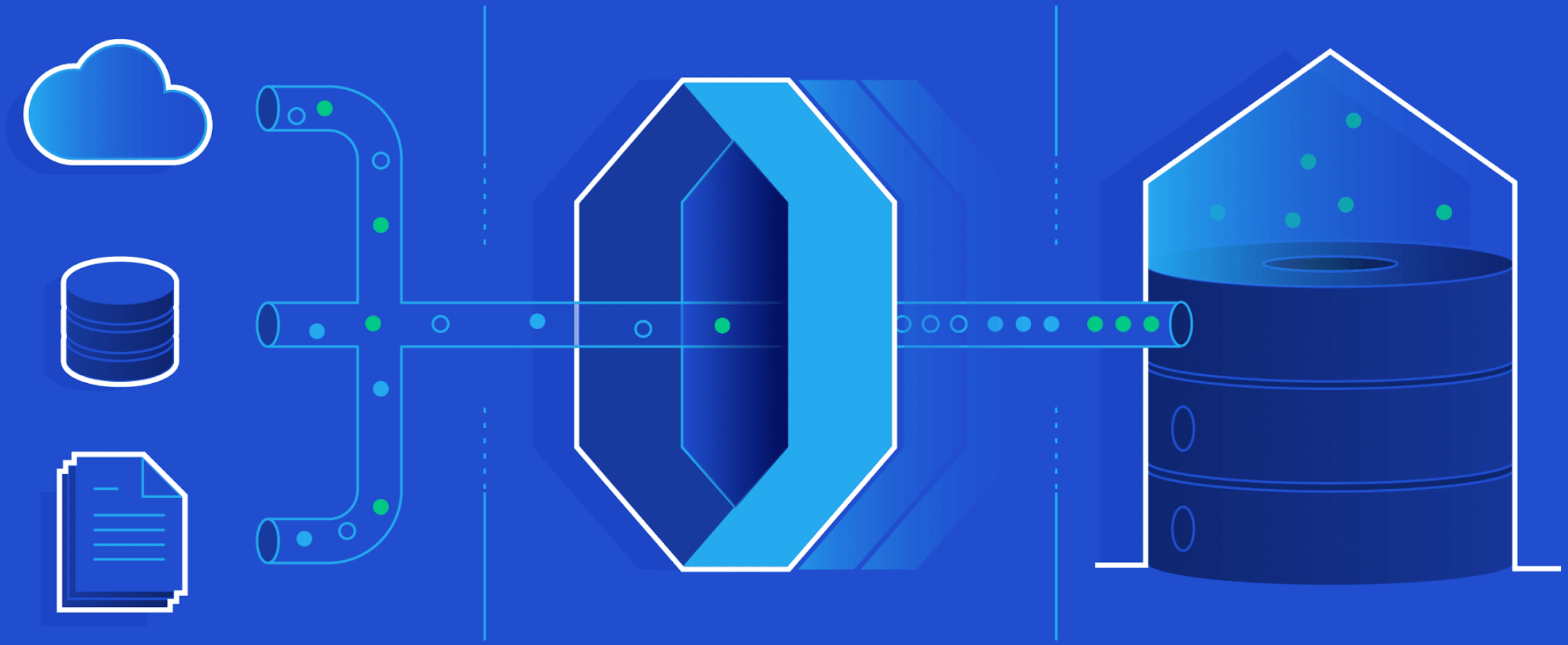
Historical Inflation and World Events

Data sets from 1913 to 2021

ESTELA, PAOLA, RAMIRO

Extract Transform Load



Extract
Files: CSV, URL

- Annual_Inflation_by_GDP_Deflator.csv
- inflation.csv
- inflation_interest_unemployment.csv
- income_growth.csv
- https://www.thebalance.com/u-s-inflation-rate-history-by-year-and-forecast-3306093

Read USA Monthly CPI Inflation CSV

```
# Read Inflation CSV Head
USA_monthly_cpi_file = "DATA/inflation.csv"
USA_monthly_cpi_df = pd.read_csv(USA_monthly_cpi_file)
USA_monthly_cpi_df.head()
```

	Unnamed: 0	Year	Jan	Feb	Mar	Apr	May	Jun	Jul	Aug	Sep	Oct	Nov	Dec
0	0	1913.0	9.8	9.8	9.8	9.8	9.7	9.8	9.9	9.9	10.0	10.0	10.1	10.0
1	1	1914.0	10.0	9.9	9.9	9.8	9.9	9.9	10.0	10.2	10.2	10.1	10.2	10.1
2	2	1915.0	10.1	10.0	9.9	10.0	10.1	10.1	10.1	10.1	10.1	10.2	10.3	10.3
3	3	1916.0	10.4	10.4	10.5	10.6	10.7	10.8	10.8	10.9	11.1	11.3	11.5	11.6
4	4	1917.0	11.7	12.0	12.0	12.6	12.8	13.0	12.8	13.0	13.3	13.5	13.5	13.7

Read Annual Inflation CSV

```
# Read Deflator CSV Head
annual_inflation_file = "DATA/Annual_Inflation_by_GDP_Deflator.csv"
annual_inflation_df = pd.read_csv(annual_inflation_file, compression='gzip')
annual_inflation_df.head()
```

	Country	Country_Code	Percent_Annual_Inflation_in_Year_1961	Percent_Annual_Inflation_in
0	Aruba	ABW	NaN	
1	Africa Eastern and Southern	AFE	1.861701	
2	Afghanistan	AFG	NaN	
3	Africa Western and Central	AFW	3.336148	
4	Angola	AGO	NaN	

5 rows × 62 columns

Extract

Annual_Inflation_by_GDP_Deflator.csv & inflation.csv

Read Unemployment CSV

```
# Read Unemployment CSV Head
Unemployment_file = "DATA/inflation_interest_unemployment.csv"
Unemployment_df = pd.read_csv(Unemployment_file)
Unemployment_df.head()
```

	country	year	Inflation, consumer prices (annual %)	Inflation, GDP deflator (annual %)	Real interest rate (%)	Deposit interest rate (%)	Lending interest rate (%)	Unemployment, total (% of total labor force) (national estimate)
0	Afghanistan	1970	NaN	NaN	NaN	NaN	NaN	NaN
1	Afghanistan	1971	NaN	NaN	NaN	NaN	NaN	NaN
2	Afghanistan	1972	NaN	NaN	NaN	NaN	NaN	NaN
3	Afghanistan	1973	NaN	NaN	NaN	NaN	NaN	NaN
4	Afghanistan	1974	NaN	NaN	NaN	NaN	NaN	NaN

Read Income CSV

```
# Read Income CSV Head
income_file = "DATA/income_growth.csv"
income_df = pd.read_csv(income_file)
income_df.head()
```

	Country Name	Country Code	Series Name	Series Code	1960	1961	1962	1963	1964	1965	...
0	Afghanistan	AFG	Adjusted net national income per capita (const...	NY.ADJ.NNTY.PC.KD	..	..	..	..	..	..	...
1	Afghanistan	AFG	Adjusted net national income per capita (annua...	NY.ADJ.NNTY.PC.KD.ZG	..	..	..	..	..	..	...
2	Albania	ALB	Adjusted net national income per capita (const...	NY.ADJ.NNTY.PC.KD	..	..	..	..	..	..	...

Extract

inflation_interest_unemployment.csv & income_growth.csv

Read U.S Inflation Rate History URL

```
# US Inflation Rate by Year From 1929 to 2023: U.S. Inflation Rate History and Forecast
url = 'https://www.thebalance.com/u-s-inflation-rate-history-by-year-and-forecast-3306093'
```

```
tables = pd.read_html(url)

US_inf_rate_hist_df= tables[0]
US_inf_rate_hist_df.head()
```

	Year	Inflation Rate YOY	Fed Funds Rate*	Business Cycle (GDP Growth)	Events Affecting Inflation
0	1929	0.6%	NaN	August peak	Market crash
1	1930	-6.4%	NaN	Contraction (-8.5%)	Smoot-Hawley
2	1931	-9.3%	NaN	Contraction (-6.4%)	Dust Bowl
3	1932	-10.3%	NaN	Contraction (-12.9%)	Hoover tax hikes
4	1933	0.8%	NaN	Contraction ended in March (-1.2%)	FDR's New Deal

Extract

https://www.thebalance.com/u-s-inflation-rate-history-by-year-and-forecast-3306093

Transform

Transform

Annual_Inflation_by_GDP_Deflator.csv & inflation.csv

Transform Annual Inflation Dataframe

```
# strip columns and only leave years
annual_inflation_df.columns = annual_inflation_df.columns.str.lstrip("Percent_Annual_Inflation_in_Year_")

#use melt function to flip information vertically
annual_inflation_transformed = annual_inflation_df.melt(id_vars=["Country", "Country_Code"],
        var_name="Year",
        value_name="Inflation_Rate")

# Filtering only columns needed
annual_inflation_transformed.columns=["country", "country_code", "year", "annual_inflation_rate"]
annual_inflation_transformed.set_index(["year", "country"], inplace=True)
annual_inflation_transformed = annual_inflation_transformed.dropna(how='any')
annual_inflation_transformed = annual_inflation_transformed.sort_values("country", ascending=True)
annual_inflation_transformed.head()
```

		country_code	annual_inflation_rate
year	country		
2004	Afghanistan	AFG	11.271432
2010	Afghanistan	AFG	3.814630
2007	Afghanistan	AFG	22.527756
2017	Afghanistan	AFG	2.403656
2011	Afghanistan	AFG	16.593347

Transform USA Monthly CPI Inflation Dataframe

```
# Selecting columns needed
USA_monthly_cpi_cols = ["Year", "Jan", "Feb", "Mar", "Apr", "May", "Jun", "Jul",
       "Aug", "Sep", "Oct", "Nov", "Dec"]
USA_monthly_cpi_transformed = USA_monthly_cpi_df[USA_monthly_cpi_cols].copy()

# Rename the column headers
USA_monthly_cpi_transformed = USA_monthly_cpi_transformed.rename(columns={"Year": "year", "Jan": "jan", "Feb": "feb", "Mar": "ma
                                                                          "Apr": "apr", "May": "may", "Jun": "jun", "Jul": "jul"
                                                                          "Aug": "aug", "Sep": "sep", "Oct": "oct", "Nov": "nov"
                                                                          "Dec": "dec"})

# Add columng for "country"
USA_monthly_cpi_transformed.loc[:, 'country'] = 'United States'

# Set "year" as type int
USA_monthly_cpi_transformed['year'] = USA_monthly_cpi_transformed['year'].astype(int)

# Set index
USA_monthly_cpi_transformed.set_index(["year","country"], inplace=True)

USA_monthly_cpi_transformed.head()
```

Transformed

inflation_interest_unemployment.csv
& income_growth.csv

Transform Unemployment Dataframe

```
# Create a filtered dataframe from specific columns
unemp_cols = ["country", "year", "Inflation, consumer prices (annual %)", "Inflation, GDP deflator (annual %)", "Unemployment, t
unemp_transformed= Unemployment_df[unemp_cols].copy()
unemp_transformed.head()

# Rename columns
unemp_transformed = unemp_transformed.rename(columns={"Inflation, consumer prices (annual %)": "cpi",
                                                      "Inflation, GDP deflator (annual %)": "gdp_deflator",
                                                      "Unemployment, total (% of total labor force) (national estimate)": "u
# Set index
unemp_transformed.set_index(["year","country"], inplace=True)

unemp_transformed.head()
```

		cpi	gdp_deflator	unemp
year	country			
1970	Afghanistan	NaN	NaN	NaN
1971	Afghanistan	NaN	NaN	NaN
1972	Afghanistan	NaN	NaN	NaN
1973	Afghanistan	NaN	NaN	NaN
1974	Afghanistan	NaN	NaN	NaN

Transform Income Dataframe

```
# Filter out income per capita in US Dollars
income_df = income_df.loc[income_df["Series Name"] == "Adjusted net national income per capita (annual % growth)",:]

# Drop unnecessary columns
income_df = income_df.drop(["Series Name","Series Code"], axis=1)

# Transposing columns to rows
cleaned_income_df = income_df.melt(id_vars=["Country Name", "Country Code"],
        var_name="Year",
        value_name="Income_Growth")

# Renaming columns
cleaned_income_df.columns=["country", "country_code", "year", "income_growth"]

# Set index
cleaned_income_df.set_index(["year","country"], inplace=True)

# Sort values
cleaned_income_df = cleaned_income_df.sort_values("country", ascending=True)

cleaned_income_df.head()
```

		country_code	income_growth
year	country		
1960	Afghanistan	AFG	..
2013	Afghanistan	AFG	..
1983	Afghanistan	AFG	..
2012	Afghanistan	AFG	..
1985	Afghanistan	AFG	..

Transform

https://www.thebalance.com/u-s-inflation-rate-history-by-year-and-forecast-3306093

Transform US Inflation Rate History Dataframe

```
# Selecting only columns needed
US_inf_rate_hist_cols = ["Year", "Business Cycle (GDP Growth)", "Events Affecting Inflation"]
US_inf_rate_hist_transformed = US_inf_rate_hist_df[US_inf_rate_hist_cols].copy()


# Rename the column headers
US_inf_rate_hist_transformed = US_inf_rate_hist_transformed.rename(columns={"Year": "year",
                                                                            "Business Cycle (GDP Growth)": "business_cycle_and
                                                                            "Events Affecting Inflation": "events_affecting_in
# Add a column for "country"
US_inf_rate_hist_transformed.loc[:, 'country'] = 'United States'


# Set index
US_inf_rate_hist_transformed.set_index(["year", "country"], inplace=True)

US_inf_rate_hist_transformed.head()
```

		business_cycle_and_gdp_growth	events_affecting_inflation
year	**country**		
1929	**United States**	August peak	Market crash
1930	**United States**	Contraction (-8.5%)	Smoot-Hawley
1931	**United States**	Contraction (-6.4%)	Dust Bowl
1932	**United States**	Contraction (-12.9%)	Hoover tax hikes
1933	**United States**	Contraction ended in March (-1.2%)	FDR's New Deal

Schema

```sql
DROP TABLE annual_inflation

CREATE TABLE annual_inflation(
    year INTEGER,
    country VARCHAR(50),
    country_code VARCHAR(5),
    annual_inflation_rate REAL
);
```

```sql
DROP TABLE us_inflation_hist

CREATE TABLE us_inflation_hist(
    year INTEGER,
    business_cycle_and_gdp_growth VARCHAR(50),
    country VARCHAR(50),
    events_affecting_inflation VARCHAR(50)
);
```

```sql
DROP TABLE usa_monthly_cpi

CREATE TABLE usa_monthly_cpi(
    year INTEGER,
    country VARCHAR(50),
    jan REAL,
    feb REAL,
    mar REAL,
    apr REAL,
    may REAL,
    jun REAL,
    jul REAL,
    aug REAL,
    sep REAL,
    oct REAL,
    nov REAL,
    dec REAL
);
```

```sql
DROP TABLE unemployment

CREATE TABLE unemployment (
    year INTEGER,
    country VARCHAR(50),
    cpi REAL,
    gdp_deflator REAL,
    unemp REAL
);
```

```sql
DROP TABLE income

CREATE TABLE income (
    year INTEGER,
    country VARCHAR(50),
    country_code VARCHAR(5),
    income_growth REAL

);
```

Load

Load

1) create_engine
2) inspector
3) pandas to load

Create database connection

```
database_name="inflation_db"
rds_connection_string = f'{protocol}://{username}:{password}@{host}:{port}/{database_name}'
engine = create_engine(rds_connection_string)
inspector = inspect(engine)
```

```
# Confirm tables
inspector.get_table_names()
```

```
['annual_inflation',
 'us_inflation_hist',
 'usa_monthly_cpi',
 'unemployment',
 'income']
```

Use pandas to load DataFrames into database

```
annual_inflation_transformed.to_sql(name='annual_inflation', con=engine, if_exists='replace', index=True)
```

```
US_inf_rate_hist_transformed.to_sql(name='us_inflation_hist', con=engine, if_exists='replace', index=True)
```

```
USA_monthly_cpi_transformed.to_sql(name='usa_monthly_cpi', con=engine, if_exists='replace', index=True)
```

```
unemp_transformed.to_sql(name='unemployment', con=engine, if_exists='replace', index=True)
```

```
cleaned_income_df.to_sql(name='income', con=engine, if_exists='replace', index=True)
```

Load

Confirm Data load

Confirm data has been added by querying the tables

```
annual_inflation =pd.read_sql_query("SELECT * FROM annual_inflation", con=engine)
annual_inflation.head()
```

	year	country	country_code	annual_inflation_rate
0	2004	Afghanistan	AFG	11.271432
1	2010	Afghanistan	AFG	3.814630
2	2007	Afghanistan	AFG	22.527756
3	2017	Afghanistan	AFG	2.403656
4	2011	Afghanistan	AFG	16.593347

```
usa_monthly_cpi=pd.read_sql_query("SELECT * FROM usa_monthly_cpi", con=engine)
usa_monthly_cpi.head()
```

	year	country	jan	feb	mar	apr	may	jun	jul	aug	sep	oct	nov	dec
0	1913	United States	9.8	9.8	9.8	9.8	9.7	9.8	9.9	9.9	10.0	10.0	10.1	10.0
1	1914	United States	10.0	9.9	9.9	9.8	9.9	9.9	10.0	10.2	10.2	10.1	10.2	10.1
2	1915	United States	10.1	10.0	9.9	10.0	10.1	10.1	10.1	10.1	10.1	10.2	10.3	10.3
3	1916	United States	10.4	10.4	10.5	10.6	10.7	10.8	10.8	10.9	11.1	11.3	11.5	11.6
4	1917	United States	11.7	12.0	12.0	12.6	12.8	13.0	12.8	13.0	13.3	13.5	13.5	13.7

```
unemployment=pd.read_sql_query("SELECT * FROM unemployment", con=engine)
unemployment.head()
```

	year	country	cpi	gdp_deflator	unemp
0	1970	Afghanistan	NaN	NaN	NaN
1	1971	Afghanistan	NaN	NaN	NaN
2	1972	Afghanistan	NaN	NaN	NaN
3	1973	Afghanistan	NaN	NaN	NaN
4	1974	Afghanistan	NaN	NaN	NaN

```
us_inflation_hist=pd.read_sql_query("SELECT * FROM us_inflation_hist", con=engine)
us_inflation_hist.head()
```

	year	country	business_cycle_and_gdp_growth	events_affecting_inflation
0	1929	United States	August peak	Market crash
1	1930	United States	Contraction (-8.5%)	Smoot-Hawley
2	1931	United States	Contraction (-6.4%)	Dust Bowl
3	1932	United States	Contraction (-12.9%)	Hoover tax hikes
4	1933	United States	Contraction ended in March (-1.2%)	FDR's New Deal

```
income=pd.read_sql_query("SELECT * FROM income", con=engine)
income.head()
```

	year	country	country_code	income_growth
0	1960	Afghanistan	AFG	..
1	2013	Afghanistan	AFG	..
2	1983	Afghanistan	AFG	..
3	2012	Afghanistan	AFG	..

Story Time

- On average how long has an inflation or deflation lasted?
- What precedent world events are similar to events happening in the world now, and how will this affect current inflation rates within the next two years in the U.S?
- Have annual income growth and employment rates kept up with the changes from annual inflation rates in the past?
- Knowing that the Federal Reserve intervenes when inflation rate is not at a healthy 2%, then by looking at annual inflation rates we can predict the Federal Reserve's next move. It will either be an Expansion to slow down inflation or a Contraction during a recession.

pgAdmin
Table Joining

```
1  SELECT annual_inflation.year,
2      annual_inflation.country,
3      annual_inflation.annual_inflation_rate,
4      income.income_growth
5  FROM income
6  RIGHT JOIN annual_inflation
7  ON income.year = annual_inflation.year
8      AND income.country = annual_inflation.country
9  ORDER BY income.country;
```

Data Output | Explain | Messages | Notifications

	year text	country text	annual_inflation_rate double precision	income_growth text
21	2012	Albania	1.0427146059999999	-0.703860789
22	2007	Albania	4.386709379	4.45343156
23	2001	Albania	3.810857369	10.97823018
24	1984	Albania	-0.028360257000000003	..
25	1998	Albania	6.730859897	6.313650126
26	1987	Albania	-0.000306093	..
27	1997	Albania	11.23964449	..

Limitations

- Time Zone (Spain)
- Format Columns to match
- Reading Large Files (Zip)
- Reformatting Date within Database

Resources

https://www.kaggle.com/datasets/neelgajare/usa-cpi-inflation-from-19132022

https://www.kaggle.com/datasets/prasertk/inflation-interest-and-unemployment-rate

https://data.world/johnsnowlabs/annual-inflation-by-gdp-deflator

https://databank.worldbank.org/source/world-development-indicators#

https://www.thebalance.com/u-s-inflation-rate-history-by-year-and-forecast-3306093

Questions