Impact of Sanctions in Venezuela

December 2023

Achintya Mishra
Joselo Pena Contreras
Mohammad Ayan Raheel
Pamela Gonzalez Muniz

Agenda

1. Executive Summary
2. The Problem Statement
3. Programs and Tools Used
4. Exploratory Data Analysis
5. Data Modeling Techniques
6. ETL Process
7. Insights
8. Analytical Model
9. Lesson Learned

Executive Summary

Executive Summary

- Venezuela has been subject to **sanctions imposed by the United States** on its **oil, external (import and export)**, and financial sectors since the last decade. This has affected the Venezuelan people, and the countries surrounding them.

- Recently the US (United States) has decided to **ease sanctions** on the Venezuelan oil sector.

- The decision to ease these sanctions could have a **significant impact** on Venezuela's economic and political landscape.

- Our aim is to **explore the macroeconomic impact** of the sanctions to derive meaningful insights for the key members involved.

The Problem

Problem Statement

- The United States has decided to ease sanctions on the Venezuelan oil sector, The decision to ease these sanctions could have a significant impact on Venezuela's economic and political landscape.

- Our aim is to **explore the macroeconomic impact of the sanctions** to derive meaningful insights for the key members involved.

Programs and Tools Used

Programming Language / Tools Used

Platforms

- MySQL Workbench
- Jupyter Notebook
- Tableau
- Microsoft Excel

Languages

- Python
- SQL

Exploratory Data Analysis

About our Data - Combining 4 Macroeconomic Datasets

- **FLAR Macro Variables Dataset**
  - FLAR (Latin American Reserve Fund) is a financial institution that provides financial stability and liquidity support to its member countries in Latin America.
  - This dataset contains economic indicators such as GDP growth, inflation rates, fiscal balances.
- **Human Development Indicators**
  - Statistics used to measure and compare the standard of living, education, and health of a country's population.
- **Trade Data**
  - It encompasses information about the import and export of goods and services.
- **Country Classification**
  - Categorization of countries based on various criteria such as income level, development status, or region.

Data Source - Links & Specifics

We have combined 4 datasets for our analysis.

- **FLAR Macro Variables Dataset**
  - Source: Latin American Reserve Fund

  → https://flar.com/en/sie-2/ — 85,080 entries

- **Human Development Indicators**
  - Source: World Bank and IMF

  → https://www.imf.org/external/datamapper/profile/VEN — 69,244 + 50 entries

- **Country & Product Trade Data**
  - Source: United States Census Bureau

  → https://www.census.gov — 357,708 entries

- **Country Classification**
  - Source: World Bank Country Classification

  → https://datahelpdesk.worldbank.org/knowledgebase/articles/906519-world-bank-country-and-lending-groups — 1,459 entries

Database: Original Data

	A	B	C	D	E	F	G	H	I	J	K	L	M	N	O
1	[indicatorID]	[NombreIndic	[Frecuencia]	[Unidad]	[NombreUnid	[Multiplicad	[Fecha]	[Fecha_Estru	[ARGENTINA]	[BOLIVIA]	[BRASIL]	[CHILE]	[COLOMBIA]	[COSTA RICA]	[REPUBLICA D
2	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Unidades	2002-03	15/03/2002	2.3989	NULL	NULL	NULL	NULL	NULL	NULL
3	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Unidades	2002-04	15/04/2002	2.8551	NULL	NULL	NULL	NULL	NULL	NULL
4	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Unidades	2002-05	15/05/2002	3.3287	NULL	NULL	NULL	NULL	NULL	NULL
5	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Unidades	2002-06	15/06/2002	3.6213	NULL	NULL	NULL	NULL	NULL	NULL
6	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Unidades	2002-07	15/07/2002	3.6071	NULL	NULL	NULL	NULL	NULL	NULL
7	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Unidades	2002-08	15/08/2002	3.6207	NULL	NULL	NULL	NULL	NULL	NULL
8	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Unidades	2002-09	15/09/2002	3.6431	NULL	NULL	NULL	NULL	NULL	NULL
9	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Unidades	2002-10	15/10/2002	3.6519	NULL	NULL	NULL	NULL	NULL	NULL
10	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Unidades	2002-11	15/11/2002	3.5256	NULL	NULL	NULL	NULL	NULL	NULL
11	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Unidades	2002-12	15/12/2002	3.4902	NULL	NULL	NULL	NULL	NULL	NULL
12	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Unidades	2003-01	15/01/2003	3.2582	NULL	NULL	NULL	NULL	NULL	NULL
13	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Unidades	2003-02	15/02/2003	3.1632	NULL	NULL	NULL	NULL	NULL	NULL
14	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Unidades	2003-03	15/03/2003	3.0747	NULL	NULL	NULL	NULL	NULL	NULL
15	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Unidades	2003-04	15/04/2003	2.8946	NULL	NULL	NULL	NULL	NULL	NULL
16	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Unidades	2003-05	15/05/2003	2.8357	NULL	NULL	NULL	NULL	NULL	NULL
17	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Unidades	2003-06	15/06/2003	2.8089	NULL	NULL	NULL	NULL	NULL	NULL
18	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Unidades	2003-07	15/07/2003	2.8013	NULL	NULL	NULL	NULL	NULL	NULL
19	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Unidades	2003-08	15/08/2003	2.9285	NULL	NULL	NULL	NULL	NULL	NULL
20	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Unidades	2003-09	15/09/2003	[illegible]	NULL	NULL	NULL	NULL	NULL	NULL

Categorical Values

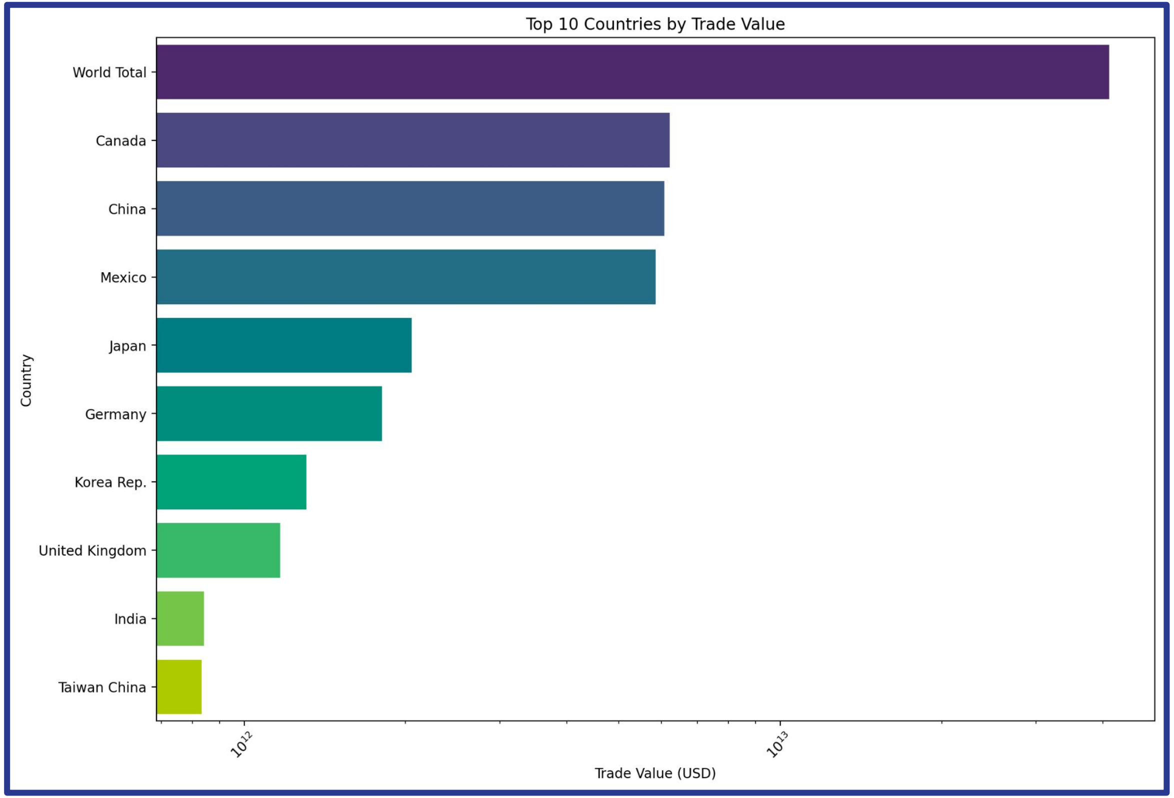
Date

Numerical Values

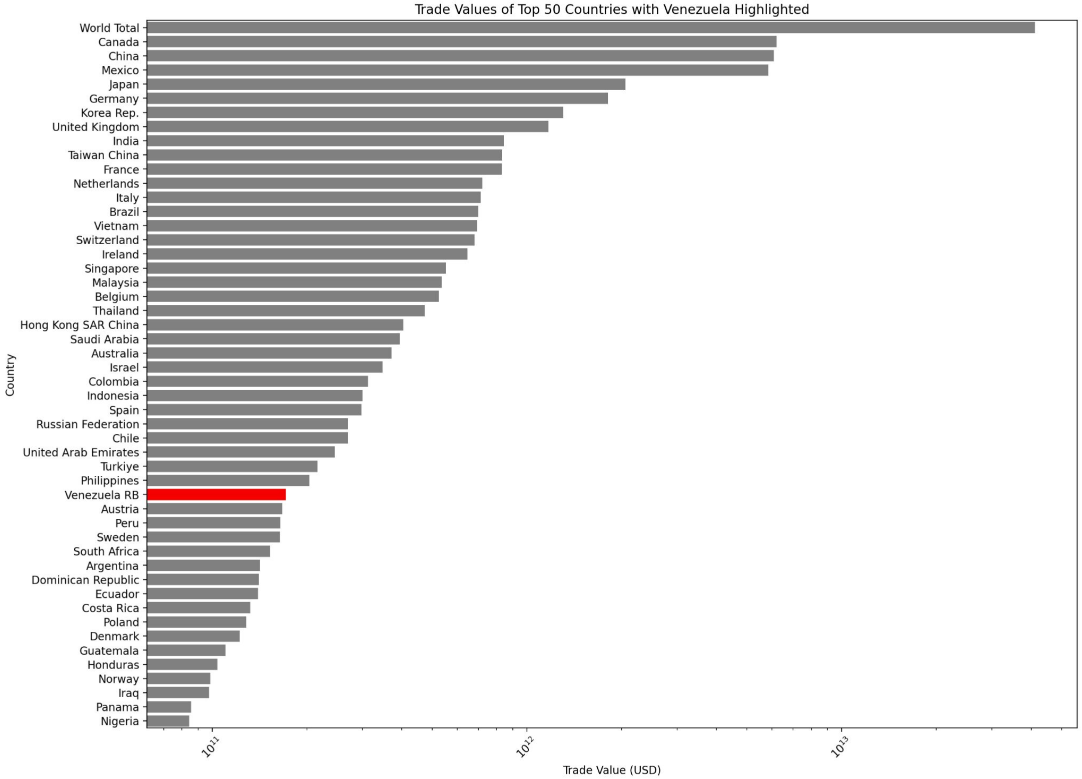
Structure of Data After Merging

Tables	Rows	Columns
Income Level	5	5
Multiplier	5	5
Region	8	8
Units	28	28
Indicator Catalogue	50	50
Country	236	236
Product	279	279
Date	1272	1272
Fact Indicator	88026	88026
Trade	357707	357707

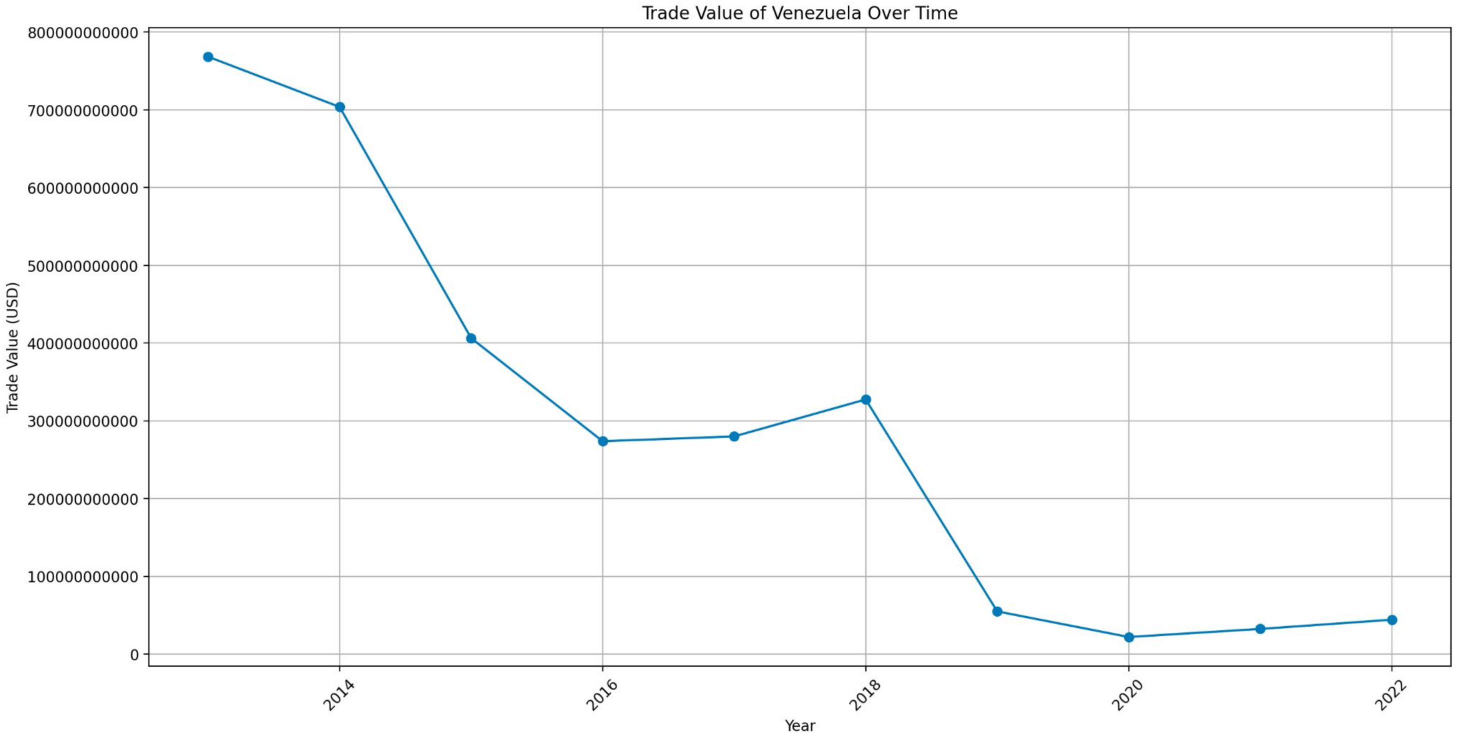
Top 10 Countries in Trade Value with the US



Venezuela is ranked 38th in terms of Trade Value



Trade Value of Venezuela Over Time



Key Characteristics of the Dataset

ACID Properties

FLAR, WB Data, IMF Data and Census Data all adhere to the ACID properties

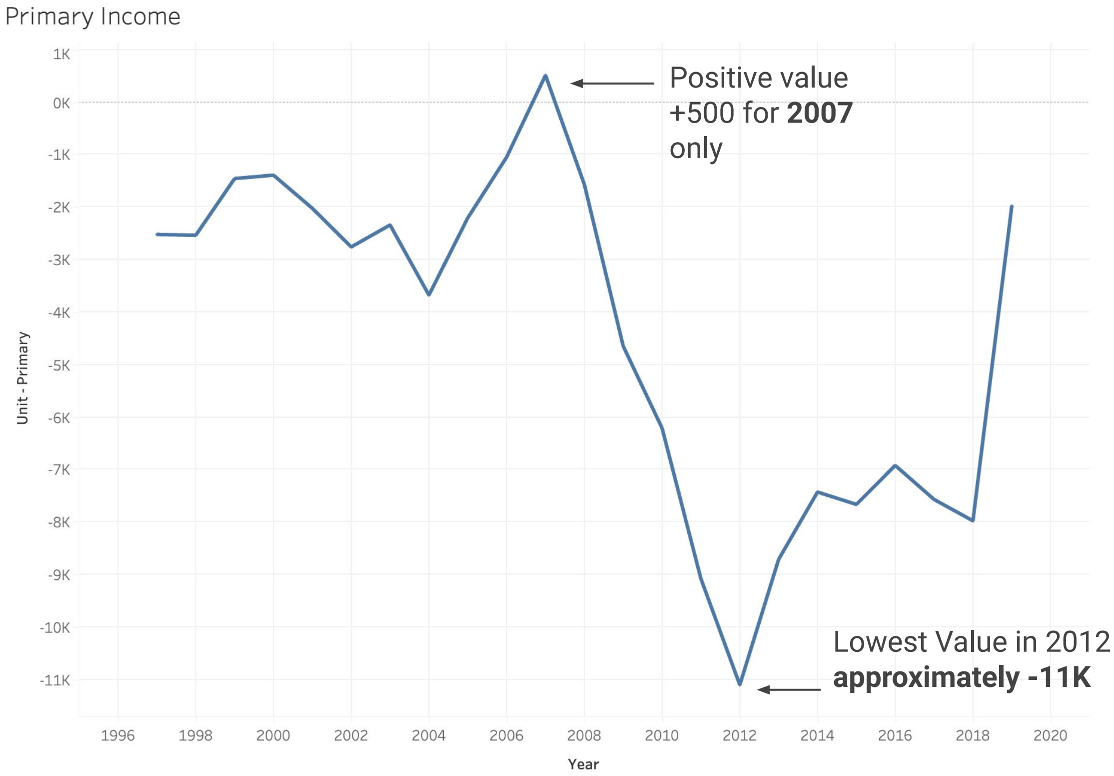
1. **Accuracy**
   No error terms to lead to discrepancies within the conclusion and recommendations
2. **Completeness**
   Data completeness achieved after merging the 4 sets - some vital information taken from all
3. **Integrity**
   Information consistency
4. **Durability**
   Relates to the performance of data overtime

+

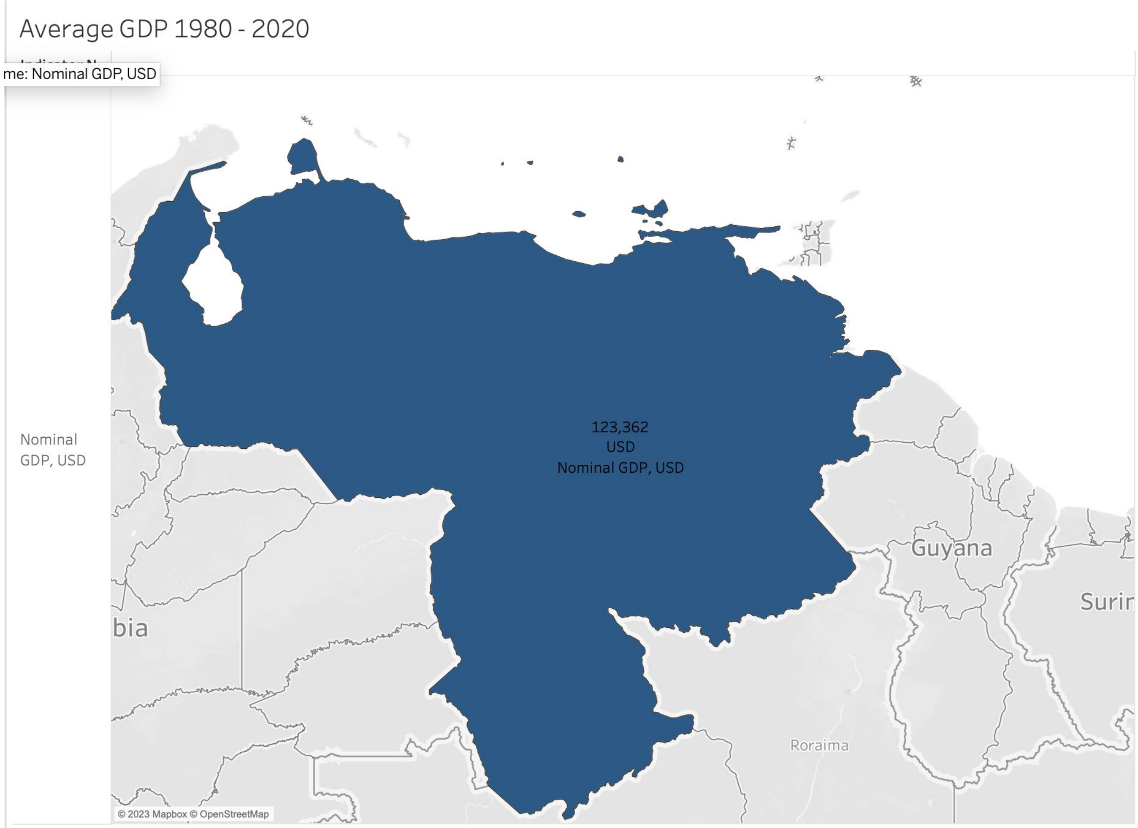
Other Properties

1. **Relevant** to our analysis
2. Data includes all the **timelines** - from 1980 till 2023
3. **Reliability -**
   Comes directly from trusted, verified sources

Primary Income Levels always below 0 except in 2007



Nominal GDP Rates on Average in USD



Average (Mean) - $ 121B from **1980 till 2023**

Minimum - $ 43.79B **2020**

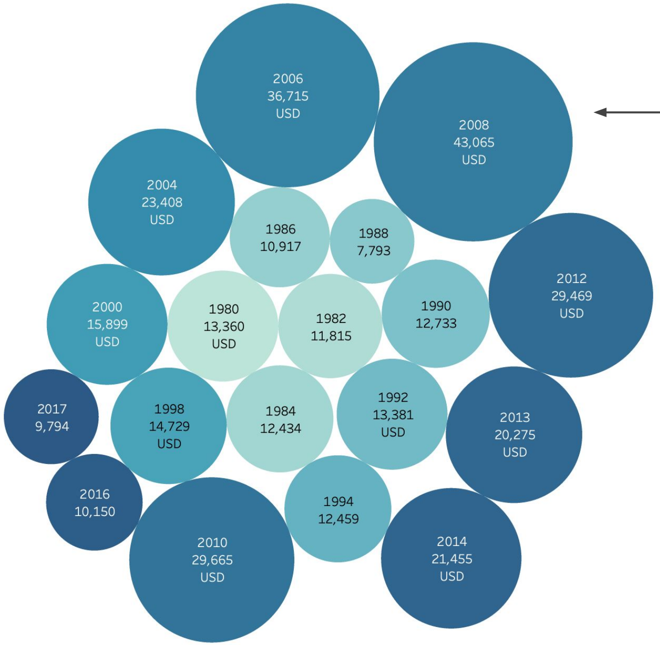
Maximum - $ 372.6B **2012**

Total Reserve per Year Trend - Increasing from 1980-2008, Decreasing from 2008-2023

General Spiraling Trend from 1980 up till 2023.

While the Reserve in absolute value was increasing at a linear rate from 1980s till 1990s, the 2000s say a wayward trend of an absolute increase but not at a rapid pace.

Total Reserves per Year



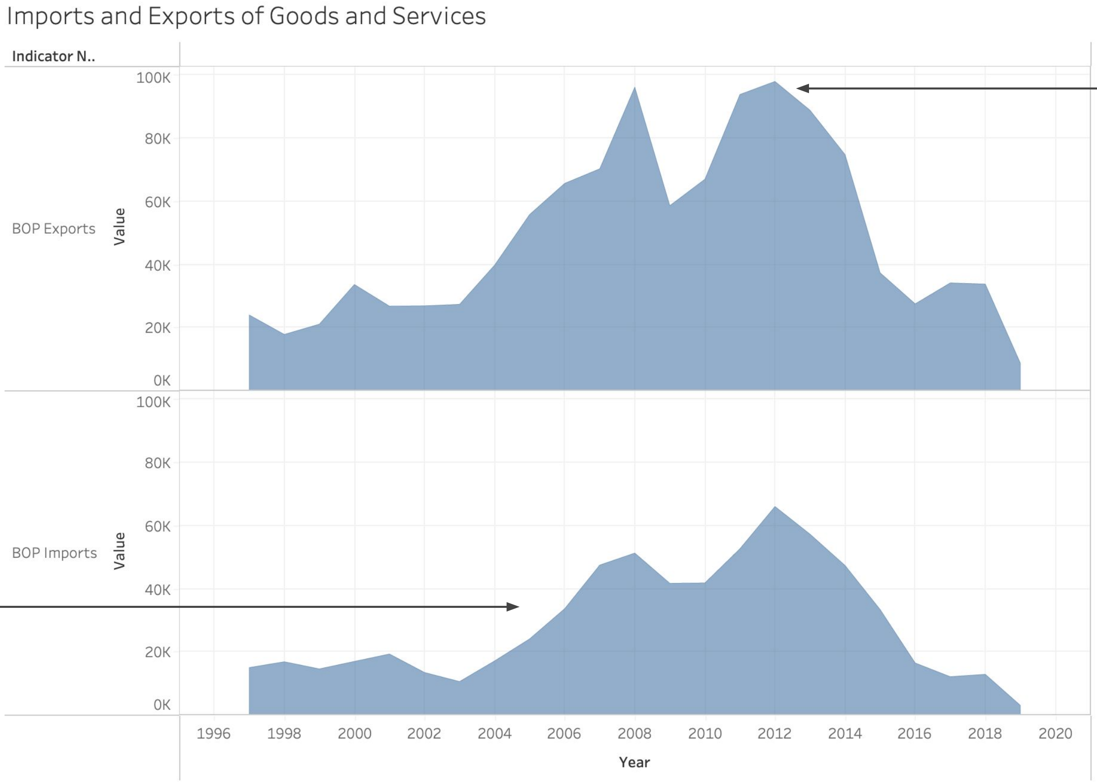
Reserves in 2008 peaked at **$43,065**

All successive years went below this value

Distribution of Imports and Exports

Similar to the Export data, BOP Imports dropped substantially with a peak in 2012 as well.

The dip in 2008 also is due to the sanctions being placed.

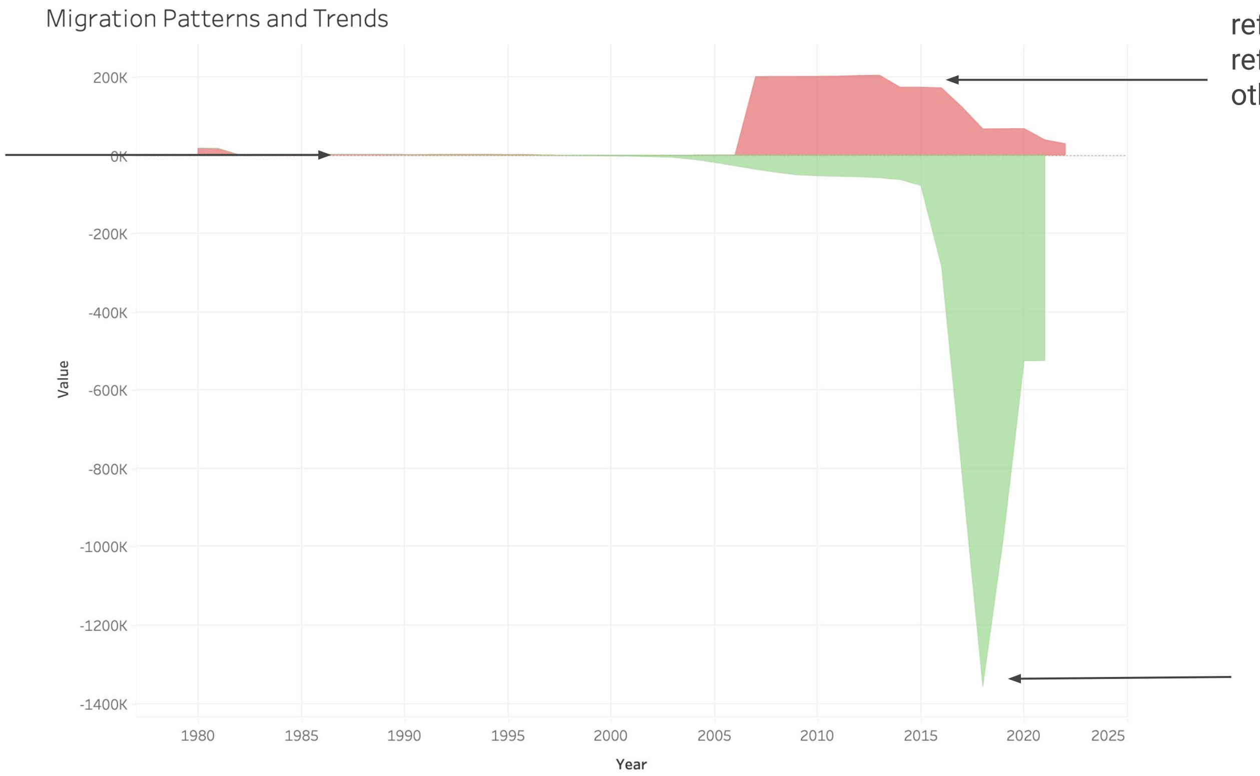


BOP Exports at their highest in 2012 - start to decrease at a rapid rate in the years that follow.

This is due to the various sanctions placed by the US on exported goods.

Migration Patterns increasing exponentially - many people migrated in 2017

Net Refugee Populations & Net Migration Patterns were both **essentially 0** until the first sanction placed in 2008.



Refugee Population reflects Venezuelan refugees living in other countries.

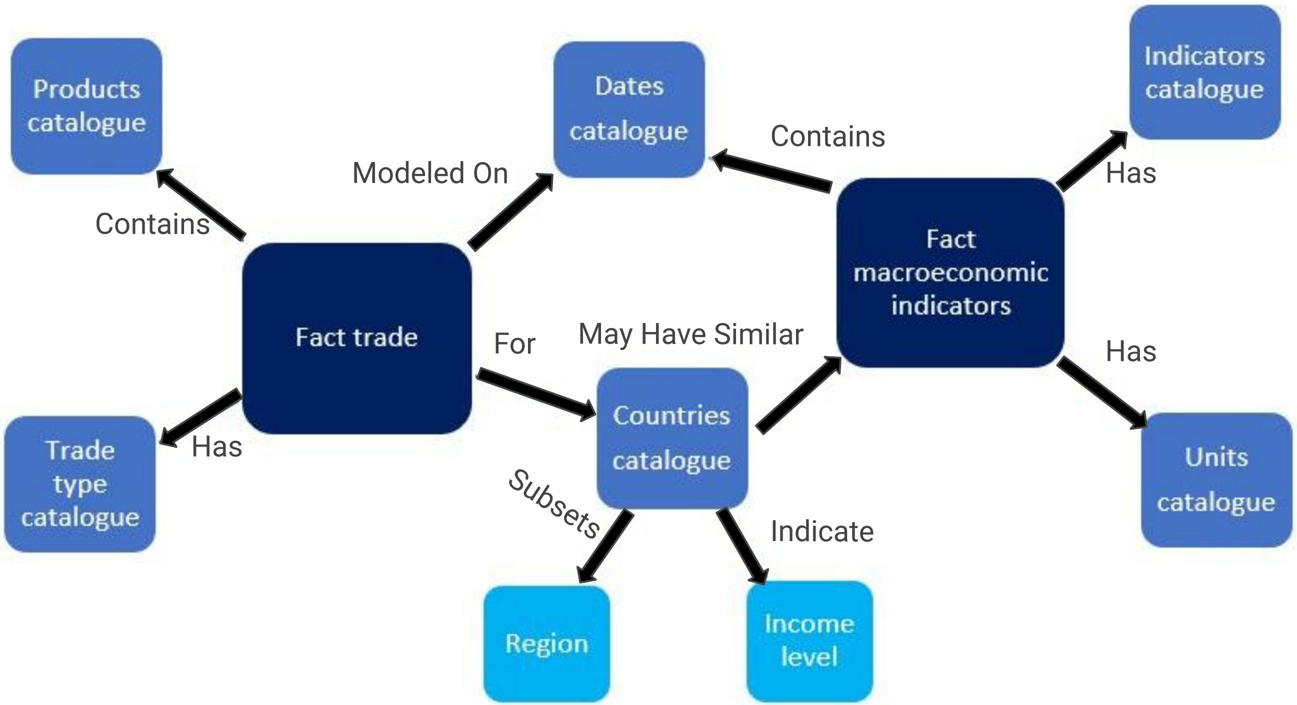
Net Migration patterns in -ve value reflects **"Brain Drain"** that is people leaving the country for opportunity elsewhere

Data Model

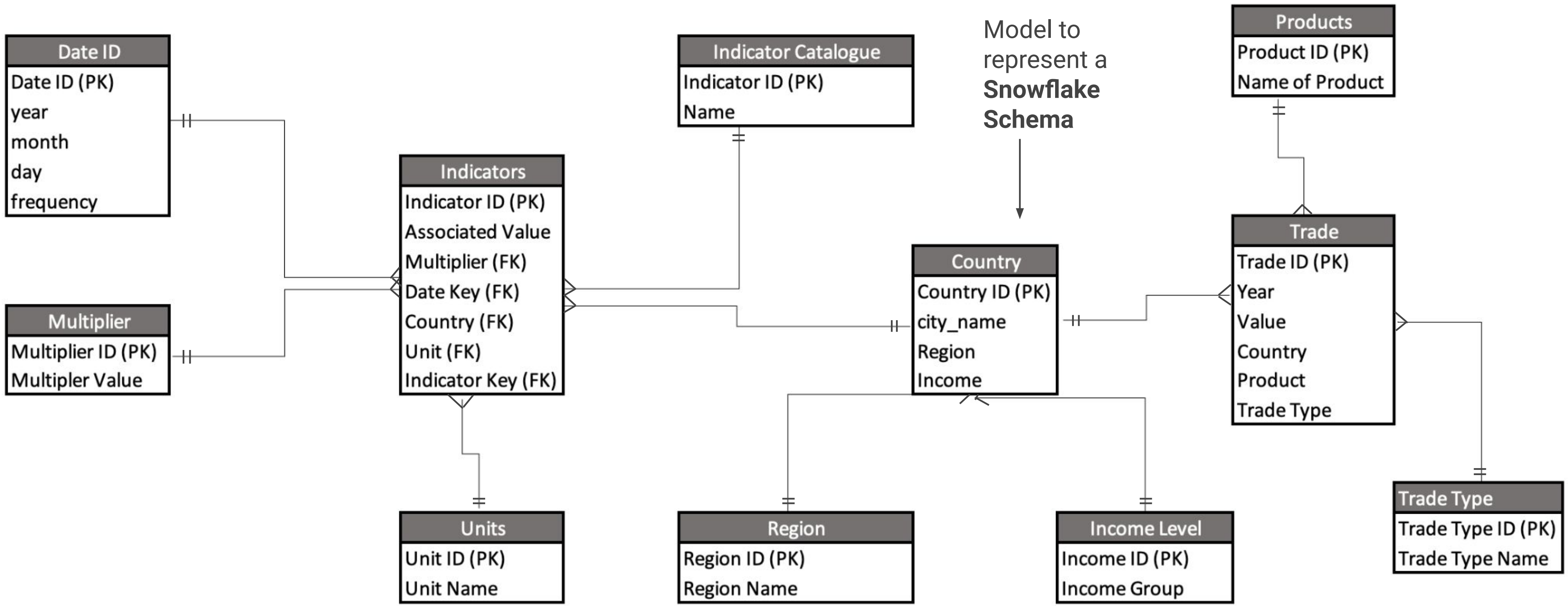
Conceptual, Logical and Physical Model Intuition

Conceptual	Logical	Physical
What exactly? Overall **Abstract View** of the entire database system Only models relationships	**What are the details?** Takes on the conceptual model and **adds structure** **Inter-entity relations**	**How to Implement?** **Final Model** Storage, Organization and access of data in DBMS

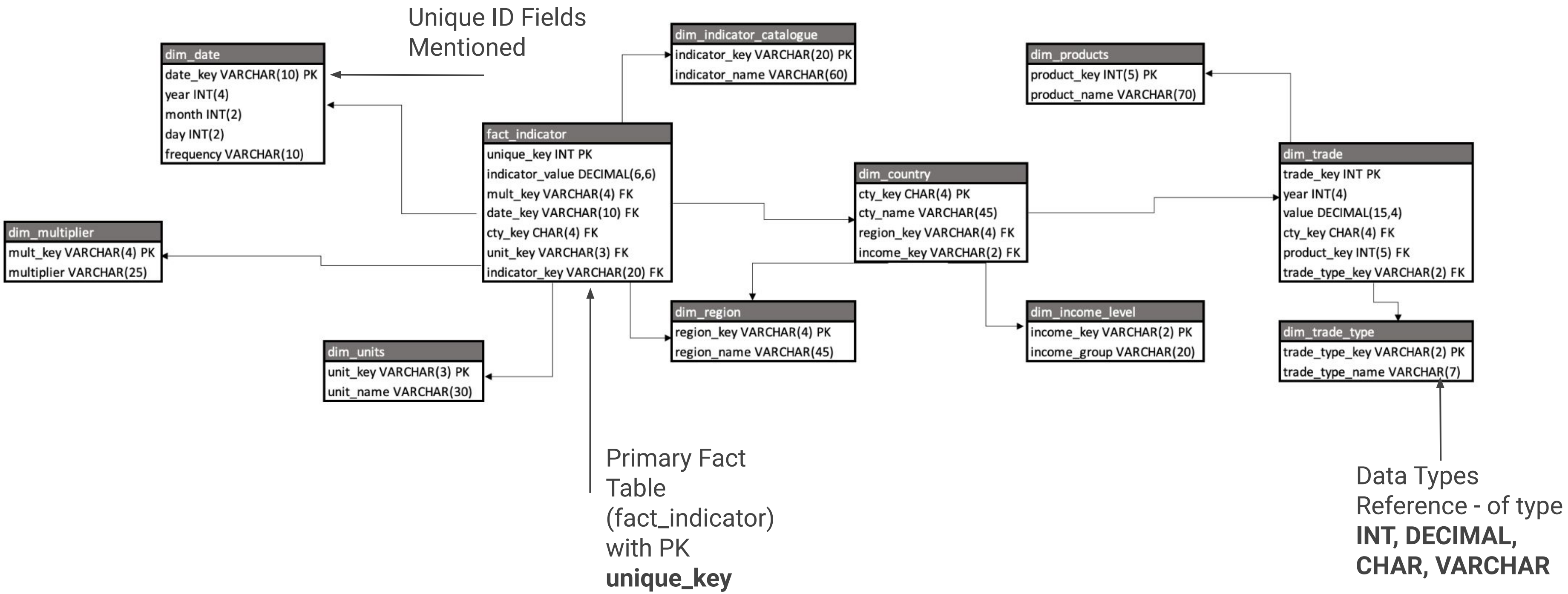
An overview of the Conceptual Model for VenezuelaDB



An overview of the Logical Model for VenezuelaDB



What the Final Physical Model Stored in DBMS Looks Like



Database Design - Utilizing a Dimensional Model

We utilized a Snowflake Schema for the modeling of the data after effectively extracting, transforming and loading the data.

Reasons for utilizing the Dimensional Model **(Kimball Approach - Flexibility + Speed)** over Relational Model -

- ❑ Dimensional Modeling is optimized for **analytical performance**
  - ❑ Better handling of data analysis requirements and query performance
  - ❑ Reporting and retrieval of aggregated data is easier
- ❑ Dimensional Modeling is engrained in **flexibility**
  - ❑ Does not include JOIN statements to undermine the performance of queries in generating analytical reports
- ❑ Dimensional Models are **adaptable**
  - ❑ New data integration with the dataset is made possible by the unnormalized/normalization to a lesser extent nature of dimensional models as compared to the relational model
- ❑ Is in line with the **business case insights** we are trying to draw
  - ❑ Using BI tools such as Tableau to model our EDA for Venezuela and SQL to create easy-to-read and easy-to-understand queries is made possible through the Dimensional Model

Steps Taken In Dimensional Modeling Snowflake Schema

Step 1

Business Process Reviewing

Transforming and merging FLAR, World Bank and IMF dataset to derive analytical insights and information.

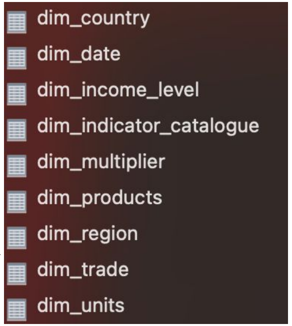
The key benefit included avoiding data redundancies duality of data

Step 2

Identifying Dimensions

Dimensions in the final filtered dataset included the following -

Country
Data
Income Level
Indicator Catalogue
Products
Region
Trade
Units



Step 3

Identifying the Fact Table

We included Indicators table as the fact table for the schema

Step 4

Determining the Granularity Level

The FLAR, IMF and WB datasets adhered to ACID properties - specifically, atomicity and completeness and were ultimately stored in our fact and dimension tables

Step 5

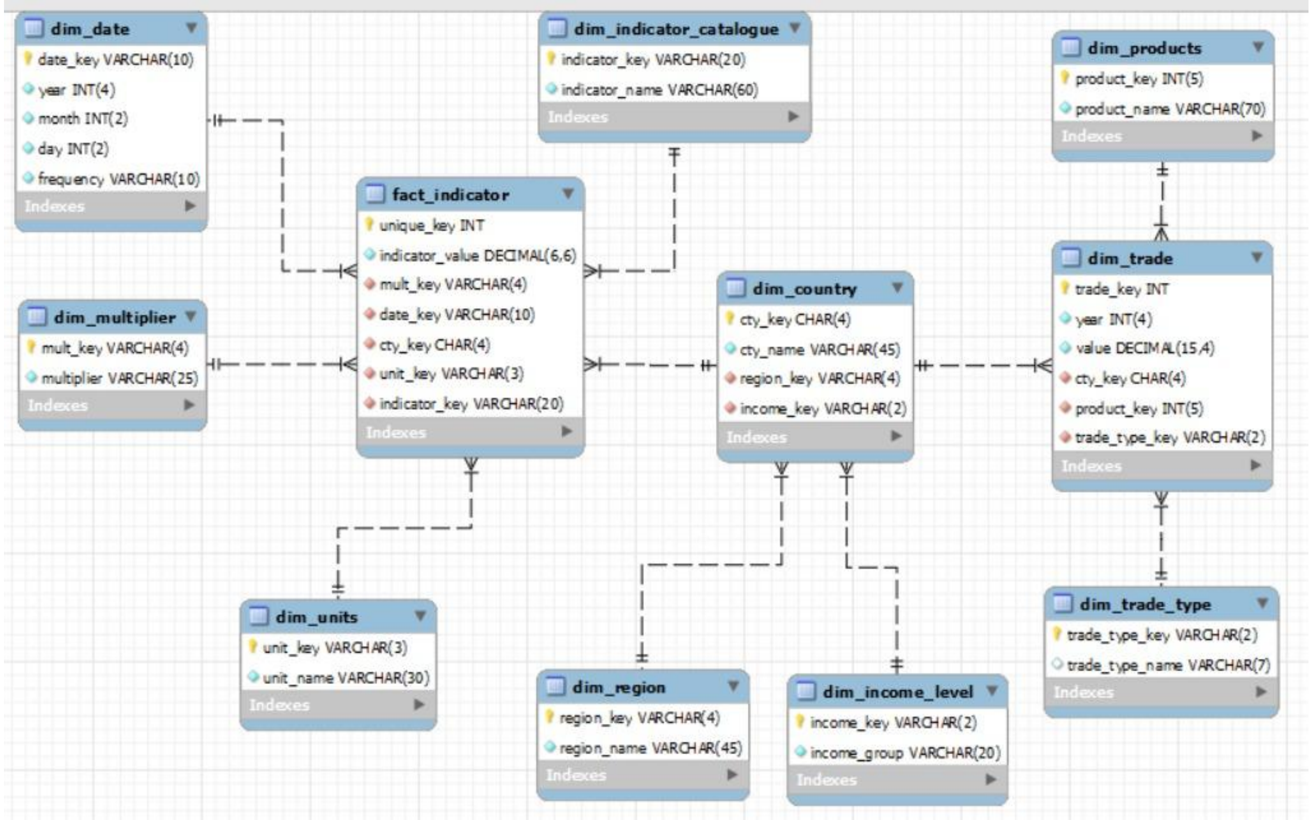
Identifying Key Relationships

The variable ‘unique_key’ a part of the ‘fact_indicator’ data was the Primary Key and each Foreign Key in the fact table was connected via a Primary Key to other dimensional tables.

ER Diagram - Snowflake Schema

The full ER diagram is used and the data is organized with analytical processing in mind.

Since optimization of results for OLTP was not a consideration, the database is considered to be for Online Analytical Processing (OLAP) and is to take shape by association with a data warehouse.



Explaining the tables and what they mean

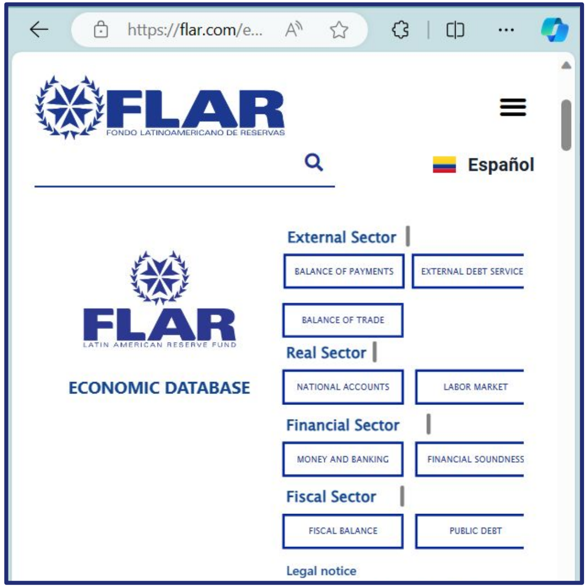
Table	Table Meaning	Included As	Cardinality w/ others	PK/FK	Types
fact_indicator	Indicator Table	Fact Table	M:1 with dim_date, dim_multipler, dim_indicator_catalogue, dim_units and dim_country	PK - unique_key FK - mult_key, FK - date_key, FK - cty_key, FK -unit_key, FK -indicator_key	INT, DECIMAL (6,6), CHAR (4) & VARCHAR (3, 4, 10)
dim_date	Date	Dimensional Table	1:M with fact_indicator	PK - date_key	VARCHAR (10), INT (2, 4)
dim_multiplier	Multiplier	Dimensional Table	1:M with fact_indicator	PK - mult_key	VARCHAR (4, 25)
dim_units	Units	Dimensional Table	1:M with fact_indicator	PK - unit_key	VARCHAR (3, 30)
dim_region	Region	Dimensional Table	1:M with fact_indicator	PK - region_key	VARCHAR (4, 45)
dim_income_level	Income Level per Country	Dimensional Table	1:M with dim_country	PK - income_key	VARCHAR (2, 20)
dim_indicator_catalogue	Indicator Catalogue	Dimensional Table	1:M with fact_indicator	PK - indicator_key	VARCHAR (20, 60)
dim_country	Country Name	2nd Level Fact Table	1:M with fact_indicator and dim_trade, M:1 dim_region, dim_income_level	PK - cty_key, FK - region_key, FK - income_key	CHAR (4) & VARCHAR (2, 4, 45)
dim_products	Products	Dimensional Table	1:M with dim_trade	PK - product_key	INT (5) & VARCHAR (70)
dim_trade	Trade	Dimensional Table	M:1 with dim_country, dim_products and dim_trade_type	PK - trade_key, FK - cty_key, FK - product_key, FK - trade_type_key	INT (4, 5), DECIMAL (15,4), CHAR (4) and VARCHAR (2)
dim_trade_type	Trade Type	Dimensional Table	1:M with dim_trade	PK - trade_type_key	VARCHAR (2,7)

Caveat - For data integrity and reduction of warning, we can potentially try to change types to INT, VARCHAR etc. in general

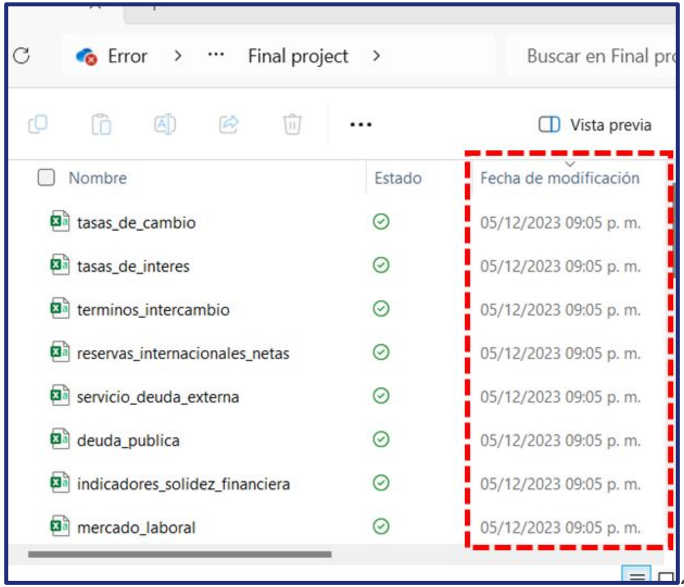
ETL Process

EXTRACT: MAIN DATA

To automate extraction we did a brief web scraping exercise with Python that obtained the urls and downloaded files in the desired path.



Web scraping and automated download



EXTRACT: MAIN DATA

ETL Extract Code

```
# SELECT DESIRED DIRECTORY
directory = "C:/Users/pamii/OneDrive/Escritorio/Master/1st quarter/Data/Final project"
if not os.path.exists(directory):
    os.makedirs(directory)

url = "https://flar.com/en/sie-2/"

response = requests.get(url)
soup = BeautifulSoup(response.content, "html.parser")

links = soup.find_all("a", href=True)
csv_links = [link["href"] for link in links if link["href"].endswith(".csv")]

for csv_link in csv_links:
    file_name = csv_link.split("/")[-1]
    file_path = os.path.join(directory, file_name)
    with open(file_path, "wb") as file:
        file_content = requests.get(csv_link).content
        file.write(file_content)
```

EXTRACT: COMPLEMENTARY DATA

International Monetary Fund

United States Census Bureau

The World Bank

We manually obtained complimentary data on trade, human development and country classifications that could help us have a broader understanding of the state of Venezuela.

TRANSFORM: MAIN DATA

We identified key transformations needed and automated them using code.

- Reshaping from wide to long
- Dropping NULL values
- Joining all indicators together
- Translating to english
- Create Primary Key

TRANSFORM: MAIN DATA

	A	B	C	D	E	F	G	H	I	J	K	L
1	[indicatorID]	[NombreIndic	[Frecuencia]	[Unidad]	[NombreUnid	[Multiplicad	[Fecha]	[Fecha_Estru	[ARGENTINA]	[BOLIVIA]	[BRASIL]	[CHILE]
2	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Unidades	2002-03	15/03/2002	2.3989	NULL	NULL	NULL
3	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Unidades	2002-04	15/04/2002	2.8551	NULL	NULL	NULL
4	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Unidades	2002-05	15/05/2002	3.3287	NULL	NULL	NULL
5	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Unidades	2002-06	15/06/2002	3.6213	NULL	NULL	NULL
6	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Unidades	2002-07	15/07/2002	3.6071	NULL	NULL	NULL
7	ENDA_X		ensual	ARS	Argentine pes	Unidades	2002-08	15/08/2002	3.6207	NULL	NULL	NULL
8	ENDA_X		ensual	ARS	Argentine pes	Unidades	2002-09	15/09/2002	3.6431	NULL	NULL	NULL
9	ENDA_X		ensual	ARS	Argentine pes	Unidades	2002-10	15/10/2002	3.6519	NULL		
10	ENDA_X		ensual	ARS	Argentine pes	Unidades	2002-11	15/11/2002	3.5256	NULL		
11	ENDA_X		ensual	ARS	Argentine pes	Unidades	2002-12	15/12/2002	3.4902	NULL		
12	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Unidades	2003-01	15/01/2003	3.2582	NULL	NULL	NULL
13	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Unidades	2003-02	15/02/2003	3.1632	NULL	NULL	NULL
14	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Unidades	2003-03	15/03/2003	3.0747	NULL	NULL	NULL
15	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Ur		/04/2003	2.8946	NULL		
16	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Ur		/05/2003	2.8357	NULL		
17	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Ur		/06/2003	2.8089	NULL		
18	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Ur		/07/2003	2.8013	NULL		
19	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Ur		/08/2003	2.9285	NULL		
20	ENDA_XDC_U	tasa de camb	Mensual	ARS	Argentine pes	Unidades	2003-09	15/09/2003	2.9209	NULL		

Spanish words

Null values

Repetitive columns

Wide data

Data was simplified and translated

unique_key	indicator_key	unit	date_code	country_code	indicator_valu	mult_code
46197	NGDP_PA_US	USD	1980	3070	69841	m_3
46198	NGDP_PA_US	USD	1981	3070	78367	m_3
46199	NGDP_PA_US	USD	1982	3070	79998	m_3
46200	NGDP_PA_US	USD	1983	3070	79672	m_3
46201	NGDP_PA_US	USD	1984	3070	57826	m_3
46202	NGDP_PA_US	USD	1985	3070	59865	m_3
46203	NGDP_PA_US	USD	1986	3070	60877	m_3
46204	NGDP_PA_US	USD	1987	3070	46854	m_3
46205	NGDP_PA_US	USD	1988	3070	60378	m_3
46206	NGDP_PA_US	USD	1989	3070	44672	m_3
46207	NGDP_PA_US	USD	1990	3070	48391	m_3
46208	NGDP_PA_US	USD	1991	3070	53382	m_3
46209	NGDP_PA_US	USD	1992	3070	60400	m_3
46210	NGDP_PA_US	USD	1993	3070	59865	m_3
46211	NGDP_PA_US	USD	1994	3070	58357	m_3

TRANSFORMATION: COMPLEMENTARY DATA

1. Reshaping trade data from wide to long.

2. Ensuring that development indicators followed the same structure as macroeconomic indicators from the main source (Adding units, multiplier, verifying same date format).

LOAD

Two key steps using MySQL:

- ➢ To create tables forward engineering our EER Model.
- ➢ To load CSV files using “LOAD DATA LOCAL INFILE”.

ETL PIPELINE

We aimed to develop an automated ETL solution that ensures ease of use and allowed scalability, reliability, and maintainability.

EXTRACT → **TRANSFORM** → **LOAD**

EXTRACT: Web scraping and automated download in Python.

TRANSFORM: Data cleaning and normalization process in Python.

LOAD: Database creation and data loading in SQL.

Insights

INSIGHTS: SANCTIONS TIMELINE

2007 — Tensions began, sanctions to specific people in Venezuela started.

2009 — Financial Crisis.

2015 — US imposes a ban on the sale of Venezuelan oil from the state owned company.

2017 — Sanction was extended to all Venezuelan companies.

INSIGHTS

Why were sanctions relevant?

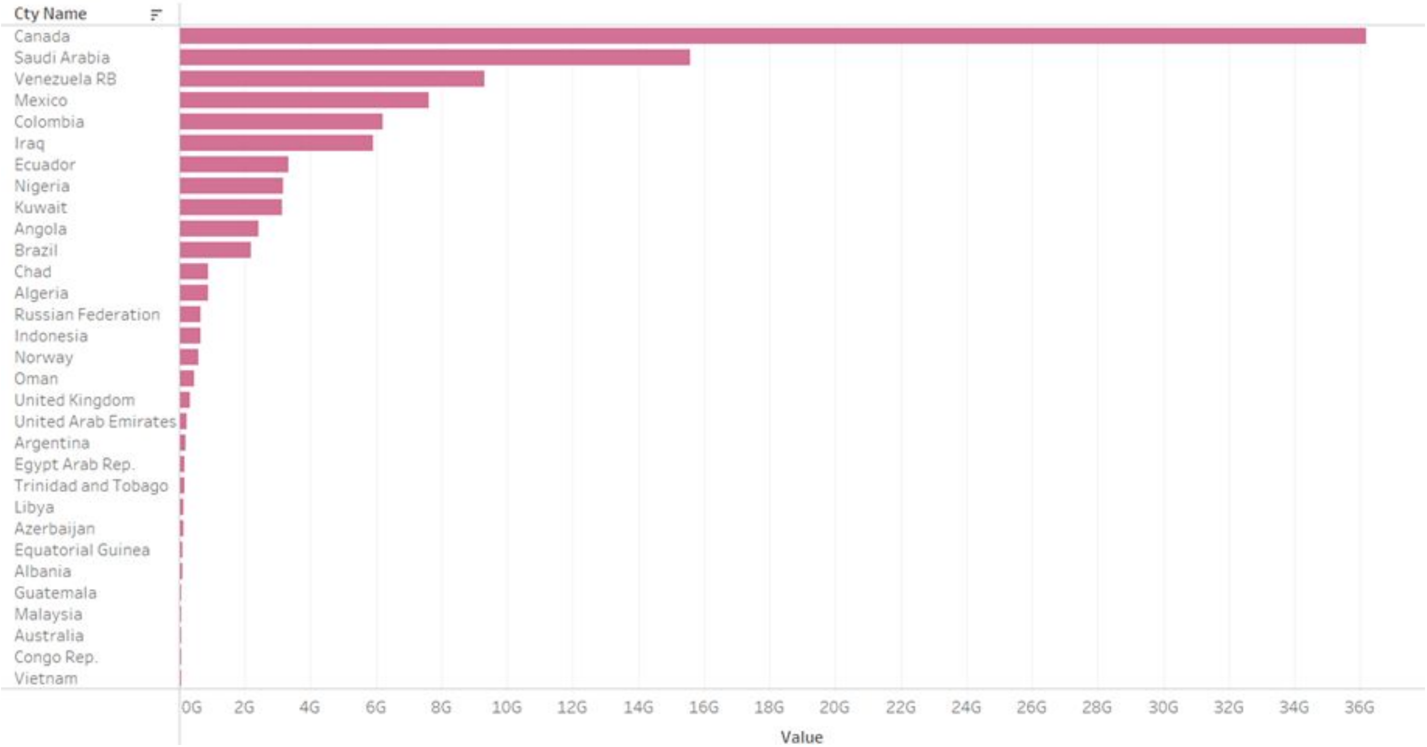
- ➢ Trade with the US represented **12% of Venezuela's GDP**.
- ➢ 85.5% of trade was frome crude oil. With other significant products being in the oil sector as well.

US imports from Venezuela in 2016

INSIGHTS

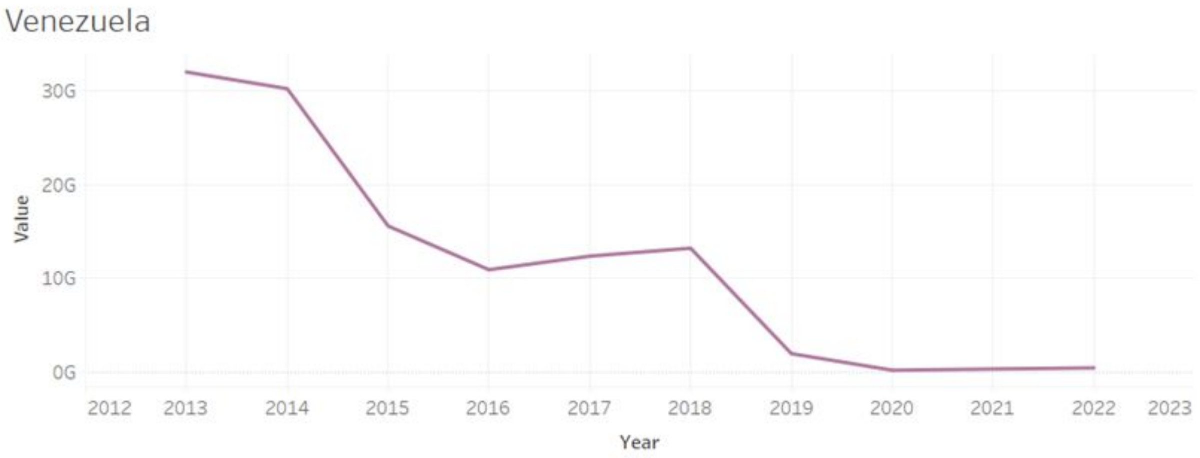
Venezuela lost its status as the **3rd biggest importer** of Crude Oil to the US.

US Imports: Top Countries for Crude Oil in 2016



INSIGHTS

While countries that had similar features to Venezuela continued to grow, the country saw a significant decline in its **commercial relationships**.



Analytical model

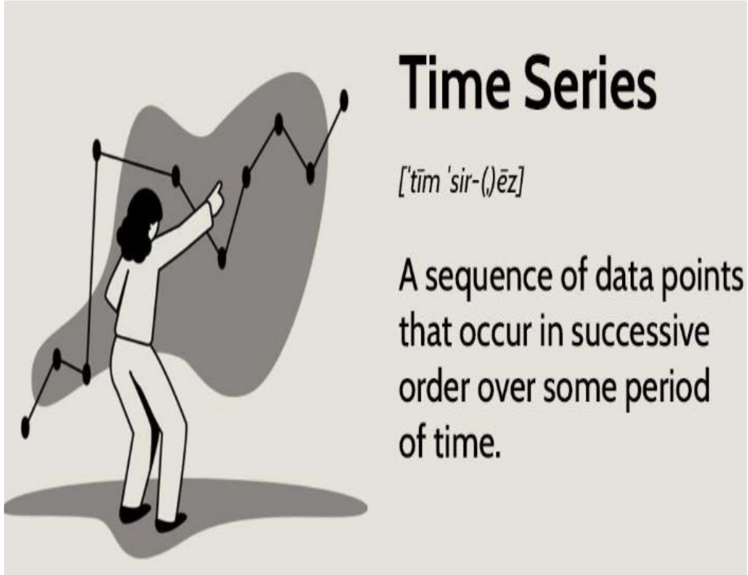
MODEL SELECTION

We are working with many **time series**

Autocorrelation: current observations are **explained by past observations**

We chose a **VARMAX** model:

- V: vector of variables
- AR: autoregressive
- MA: moving average
- **X: eXogenous variables**



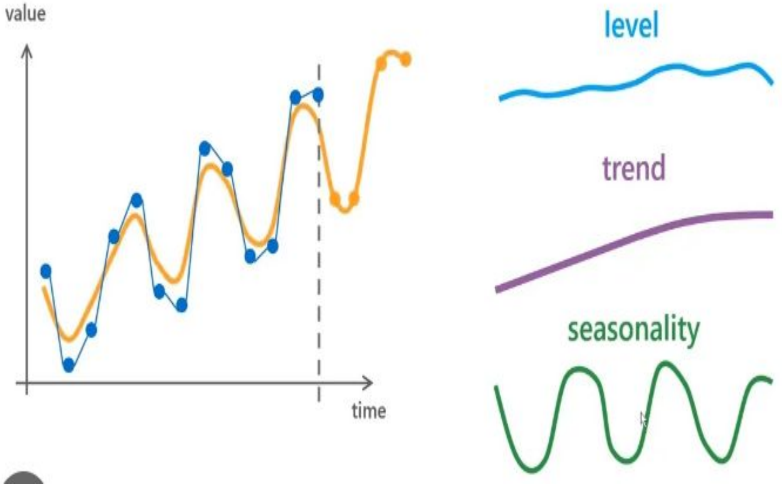
VARIABLES ADJUSTMENTS

Stationarity: variables should not have an explosive trend so the betas and forecast can be interpreted

We modified the series calculating either the difference or the percent change to make them stationary

Trend: VARMAX model can fit a trend

Seasonality: this component was removed



MODELS IMPLEMENTATION

Goal: measure the effect of sanctions on Venezuela. They are represented in dummy variables

We fit two models: one for quarterly variables and other for yearly variables

Macroeconomic variables are quarterly. Development variables are yearly

Model 1 (quarterly):

VARMAX(8,3)

```
'BOP Capital and Financial Account', 'BOP Current Account', 'Exports', 'External Debt', 'External Debt Services', 'Real GDP', 'Foreign Exchange Reserves'
```

Model 2 (yearly):

VARMAX(5,1)

```
'Mortality rate, under-5 (per 1,000 live births)', 'Net migration', 'Number of infant deaths', 'Refugee population by country or territory of asylum'
```

MODELS RESULTS

- ➢ The partial restrictions on the oil sector had a big impact on **3 key macroeconomic variables**
- ➢ Exports decreased by **24%**
- ➢ Debt service decreased by **23%**
- ➢ GDP fell by **6%**

	Exports	External Debt Services	Real GDP
2008Q4 Financial crisis	-71%	-1%	1%
2009Q1 Financial crisis	5%	-70%	-4%
2015Q1 Partial Restrictions on Oil	-24%	-23%	-6%

MODELS RESULTS

- ➢ Mortality rate, biggest impact from the sanctions of 2016, estimated increase in the mortality rate of **4.95 per 1,000 live births.**
- ➢ Infant deaths, estimated total effect of **2,848 infant deaths** related to sanctions of 2016.
- ➢ Net Migration, estimated increase of **565,400 people migrating out** of Venezuela associated to sanctions of 2018.

	Child Mortality Rate	Infant Deaths	Net Migration
2007 Tensions began	-0.69	46	-22,870
2009 Financial Crisis	0.11	183	-22,950
2016 Partial Restrictions on Oil	4.95	2,848	-230,000
2018 Full Restrictions on Oil	-0.22	-222	-565,400

Lessons Learned

CONCLUSION

- ➢ There are repercussions of the various sanctions that were felt by Venezuela in terms of their **socio-economic outcomes**

- ➢ These **not only affect the macroeconomic and trade variables but the human development indexes** and socioeconomic factors that contribute to the overall living conditions in the country

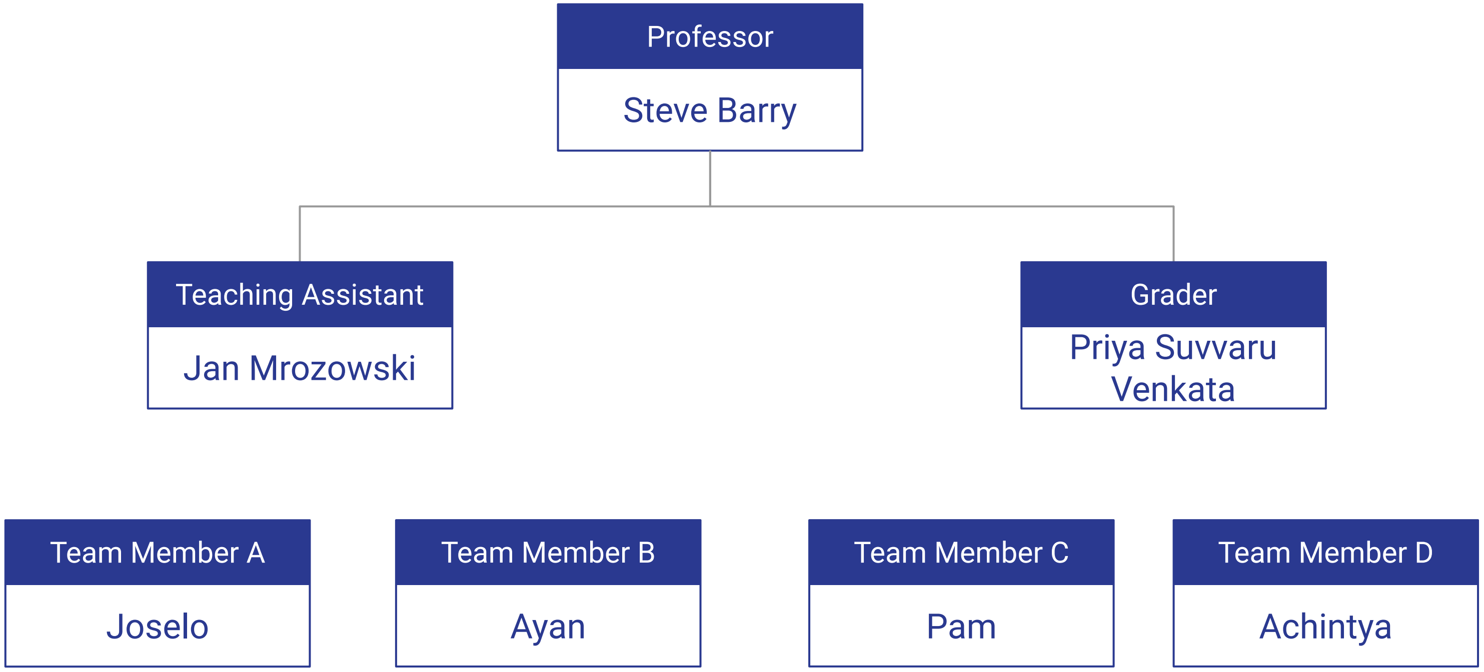
- ➢ Had the sanctions not been placed, the country's socio-economic outlook could be significantly different considering the **natural resources** the country has to **extract and export** it to other countries

- ➢ Since sanctions were **lifted this year (2023)**, it is important to **revisit the data** in the following 4/5 years to evaluate the economic and socio-demographic improvement

- ➢ Is not only a couple of variables that determine the development and economic growth of a country, but the **interaction between them on a macroeconomic level**

Appendix

The Team


Professor
Steve Barry
Teaching Assistant
Jan Mrozowski
Grader
Priya Suvvaru Venkata
Team Member A
Joselo
Team Member B
Ayan
Team Member C
Pam
Team Member D
Achintya

References

Conceptual, Logical and Physical Model -
https://www.thoughtspot.com/data-trends/data-modeling/conceptual-vs-logical-vs-physical-data-models

https://www.couchbase.com/blog/conceptual-physical-logical-data-models/

Dimensional Modeling Requirements -
https://www.astera.com/knowledge-center/dimensional-modeling-guide/

Code

Exploratory Data Analysis

```
In [1]: #pip install --upgrade mysql-connector-python

In [3]: import mysql.connector

        # Establishing a connection to the vendb database
        config = {
            "host": "127.0.0.1",
            "user": "root",
            "password": "RootRoot",
            "database": "vendb"
        }
        cnx = mysql.connector.connect(**config)
        cursor = cnx.cursor()

In [4]: cursor.execute("SHOW TABLES;")
        tables = cursor.fetchall()
        print("Tables in vendb database:")
        for table in tables:
            print(table[0])

        Tables in vendb database:
        dim_country
        dim_date
        dim_income_level
        dim_indicator_catalogue
        dim_multiplier
        dim_products
        dim_region
        dim_trade
        dim_units
        fact_indicator
```

Code

Exploratory Data Analysis

```python
# Analyzing the trade data to provide more trends

# Loading Trade Data
trade_data = pd.read_csv('trade.csv')

# Converting 'year' to datetime
trade_data['year'] = pd.to_datetime(trade_data['year'], format='%Y')
trade_data.set_index('year', inplace=True)

# Grouping by year and sum the trade values
trade_values_by_year = trade_data['value'].resample('Y').sum()

# Calculating the percentage change in trade values from the previous year
trade_values_by_year_pct_change = trade_values_by_year.pct_change()

# Ploting the percentage change in trade values over time
plt.figure(figsize=(14, 7))
plt.plot(trade_values_by_year_pct_change.index.year, trade_values_by_year_pct_change.values, marker='o')
plt.title('Percentage Change in Trade Values Over Time')
plt.xlabel('Year')
plt.ylabel('Percentage Change')
plt.grid(True)
plt.show()
```

Code

Exploratory Data Analysis

```python
# Sorting the countries by trade value.
trade_values_sorted = trade_values_sum.sort_values(by='value', ascending=False)

# Creating a bar plot.
plt.figure(figsize=(14, 10))

# Highlighting Venezuela in red
highlight = ['red' if country == 'Venezuela RB' else 'grey' for country in trade_values_sorted['cty_name']]

sns.barplot(x='value', y='cty_name', data=trade_values_sorted.head(50), palette=highlight)
plt.title('Trade Values of Top 50 Countries with Venezuela Highlighted')
plt.xlabel('Trade Value (USD)')
plt.ylabel('Country')
plt.xscale('log')
plt.tight_layout()

# Show the plot
plt.show()
```

Code

Exploratory Data Analysis

```python
import pandas as pd

# Load the dates data to merge with the trade data
dates_df = pd.read_csv('dates.csv')

# Now let's try to merge the trade data with the dates data again to get the trade values over time for Venezuela
venezuela_trade_over_time = trade_df.merge(dates_df, on='date', how='left')
venezuela_trade_over_time = venezuela_trade_over_time[venezuela_trade_over_time['cty_key'] == 3070]  # Filter for Ve

# Group by year and sum the trade values
venezuela_trade_over_time = venezuela_trade_over_time.groupby('year')['value'].sum().reset_index()

# Plot the trade value of Venezuela over time
plt.figure(figsize=(14, 7))
plt.plot(venezuela_trade_over_time['year'], venezuela_trade_over_time['value'], marker='o')
plt.title('Trade Value of Venezuela Over Time')
plt.xlabel('Year')
plt.ylabel('Trade Value (USD)')
plt.grid(True)
plt.tight_layout()
```

Code

ETL Load Code

```sql
-- -----------------------------------------------------
-- Table `vendb`.`dim_units`
-- -----------------------------------------------------
CREATE TABLE IF NOT EXISTS `vendb`.`dim_units` (
  `unit_key` VARCHAR(3) NOT NULL,
  `unit_name` VARCHAR(30) NOT NULL,
  PRIMARY KEY (`unit_key`))
ENGINE = InnoDB;
```

```sql
LOAD DATA LOCAL INFILE 'C:/Users/pamii/OneDrive/Escritorio/Master/1st quarter/Data/Final project/SQL Files/units.csv'
INTO TABLE dim_units
FIELDS TERMINATED BY ','
ENCLOSED BY '"'
LINES TERMINATED BY '\n'
IGNORE 1 ROWS;
```