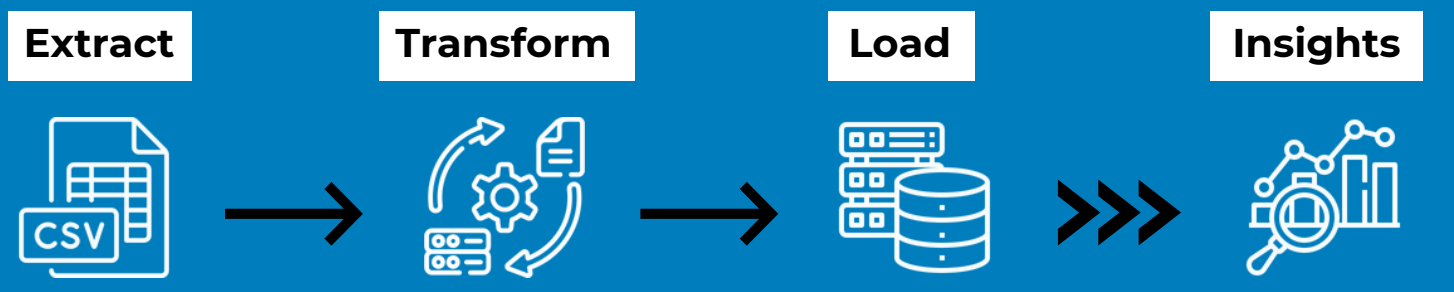
Global Food Prices

Feeding the World: Global Food Prices Data Warehouse Report

ETL Data Pipeline



Created by: Abdelrahman Alarqan

Created at: June 5, 2023

Table of Contents

CONTENTS

INTRODUCTION

What is the World Food Program

The leading humanitarian organization saving lives and changing lives, delivering food assistance in emergencies and working with communities to improve nutrition and build resilience. www.wfp.org

Context:

Global food price fluctuations can cause famine and large population shifts. Price changes are increasingly critical to policymakers as global warming threatens to destabilize the food supply.

Inspiration:

This data would be particularly interesting to pair with currency fluctuations, weather patterns, and/or refugee movements.

Data Understanding 01

Over 740k rows of prices obtained in developing world markets for various goods. Data includes information on the country, market, price of goods in local currency, quantity of goods, and month recorded.

#	Column	Description
1	Country Name	The Name of the country
2	Market Name	The name of the market
3	CM Name	The name of the commodity
4	Currency Name	The name of the currency
5	Measure Name	The unit of goods measurement (commodity)
6	MType Name	The name of the market type
7	Month	The month recorded
8	Year	The year recorded
9	Price	The paid price
10	Commod Scource	The Source supplying price information

Data Understanding 02

Always the data need to clean and transform such as:

- Add new columns
- Data correction
- Get third-party's data such as USD Price

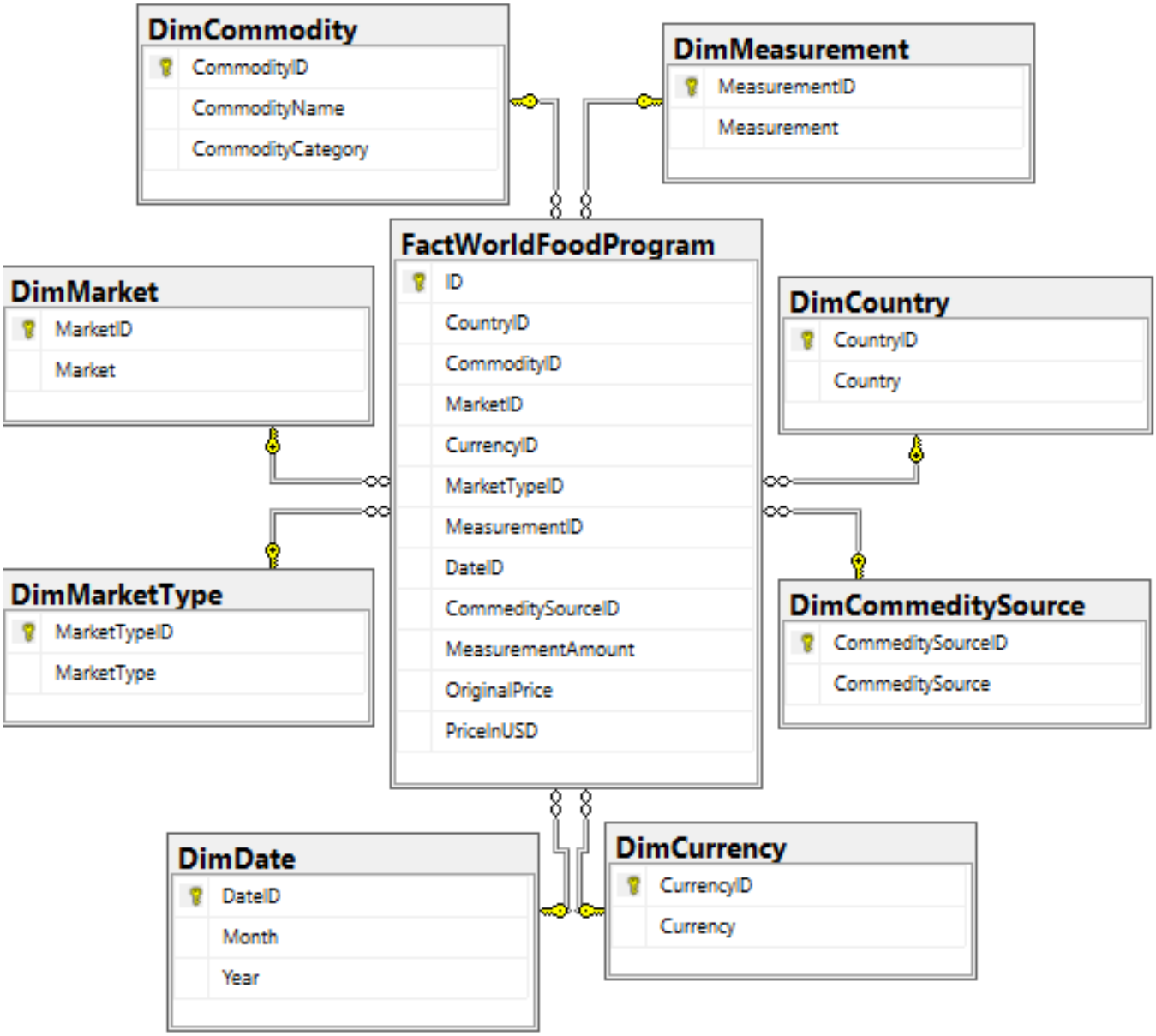
Data Cleaning & Transformation:

- The commodity source has many duplicated data so we need to correct them.
- The price is in multiple currencies so we need to add a new column and specify one currency to get a clear analysis, so I got an API with the fresh price of USD for each currency.
- The commodity name has the category and the name in one column, so I split the column into two new columns (Commodity Name, Commodity Category).
- Measurements look like these: 5KG, 10KG or 500G, so I have created a new column called measurement amount and I extracted only the amount to be like this: 5,10,500 and the old column contain the measurement only like these: KG, G, Pound.
- At the end - data type correction.

I used Python to clean the data before the ETL Process because I used an API while transforming the data, and we can transform it by SSIS while the ETL Process.

Set Up the Data Model

The data model that I created was **Star Schema,** as we see here there are fact and dimension tables for the schema.



Setup Data Warehouse (SSMS)

After we created the data model we need to apply the model to the database so I choose SQL Server Management Studio for this project.

#	Table	Description
1	Staging Area Table	This table is the landing zone of the data sources.
2	Fact Table	Contains the sorted countries with ID
3	Dim Country	Contains the sorted countries with ID
4	Dim Commodity	Contains the commodity name and category with ID
5	Dim Commodity Source	Contains the commodity source name and ID
6	Dim Date	Contains the sorted months and years with ID
7	Dim Market	Contains the market name and type with ID
8	Dim Currency	Contains the currency name with ID
9	Dim Measurements	Contains the measurement name with ID

Set Up the ETL Process (SSIS)

01

Extract

- We will extract the data from the CSV file to the staging area to be landed there.
- Then we will move and update the data into dimension and fact tables.

02

Transform

- The data now are in the staging area, so we will transform and move it into the dimensions and facts tables.
- How to do that? simply we will get the distinct values for each dimension and store them in the dimension table then we will update the ID value in the staging with the same column ID in the dimension.

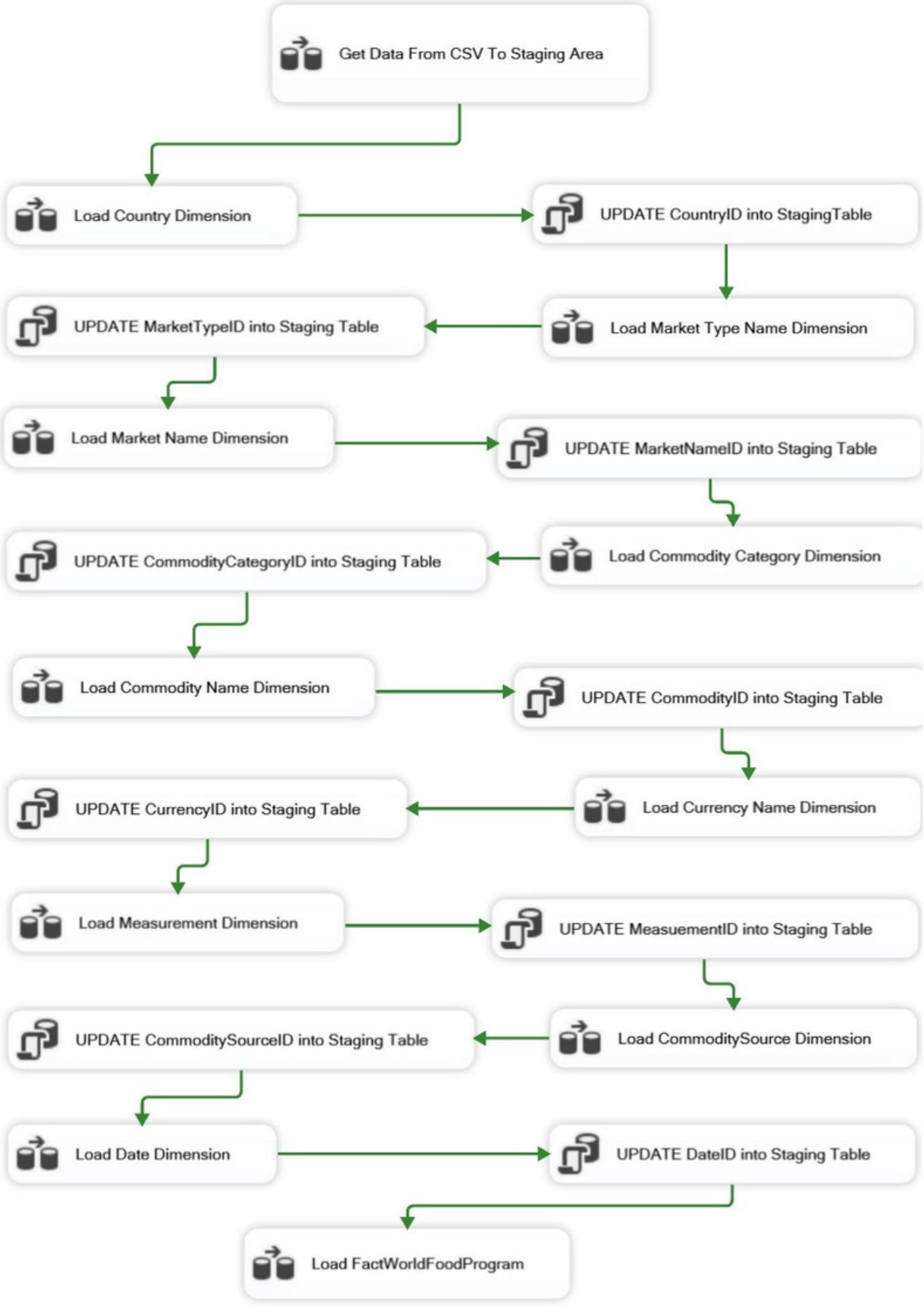
03

Load

Finally, we will move the data from the staging table into the fact table with no categorical values only IDs and measurements

On the next page, I attached an image of the ETL workFlow in exported from SSIS that you can check and see.

Set Up the ETL Process (SSIS)

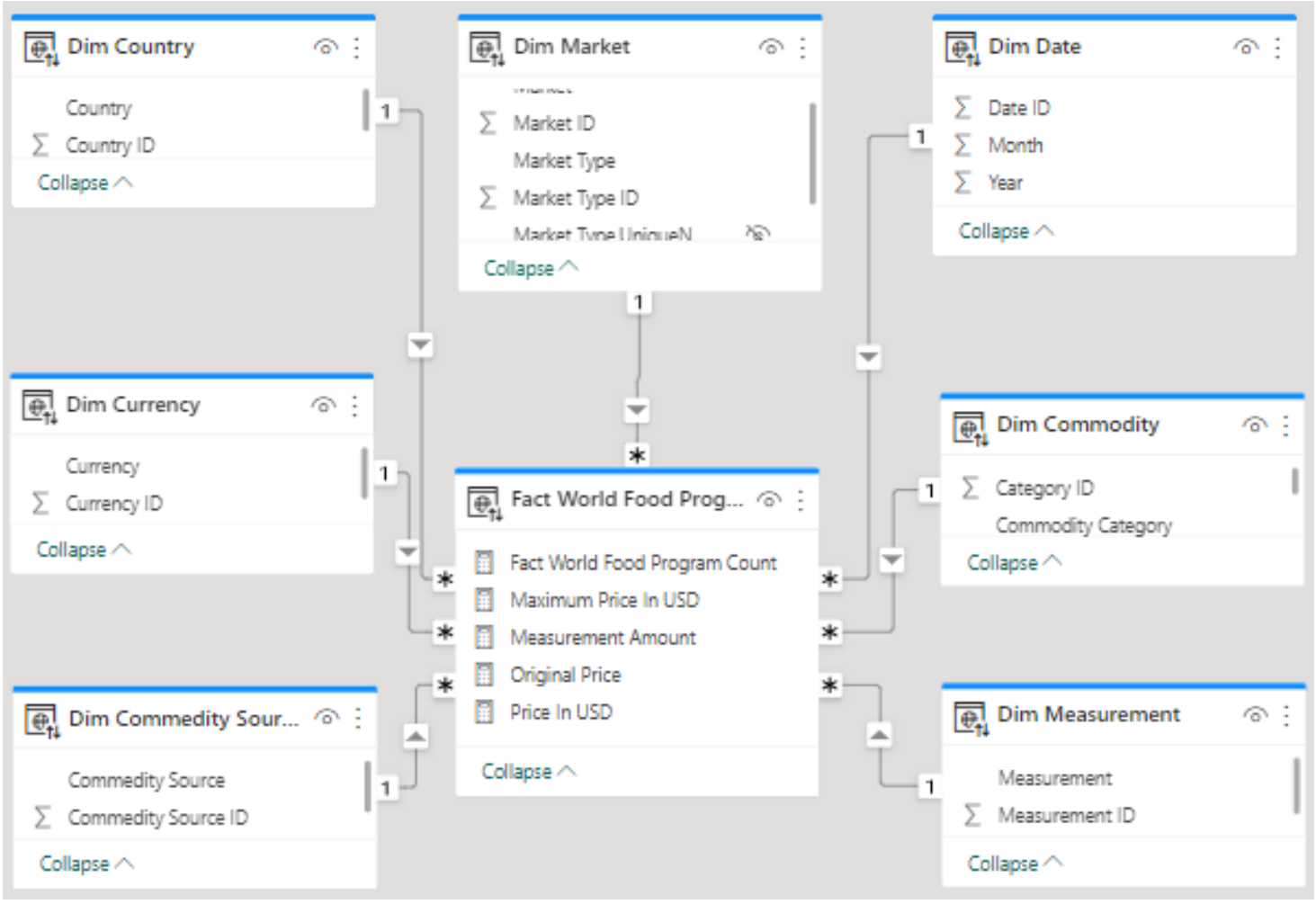


Set Up Analysis Cube (SSAS)

Here is our cube that contains five measures:

- Original Price
- Price in USD
- Maximum Price in USD
- Count of rows
- Measurement Amount

Also contains seven main dimensions as the image attached



Set Up Report (Power BI) 01

Let's start our data storytelling

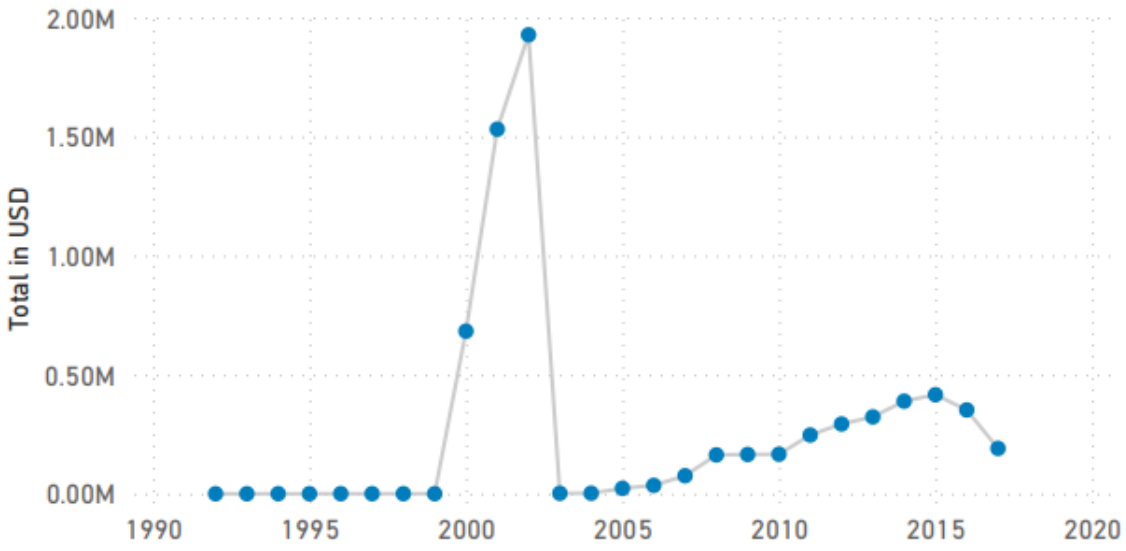
The number of expenses

743.91K

Total expenditures from 1992 to 2017 (USD)

6.99M

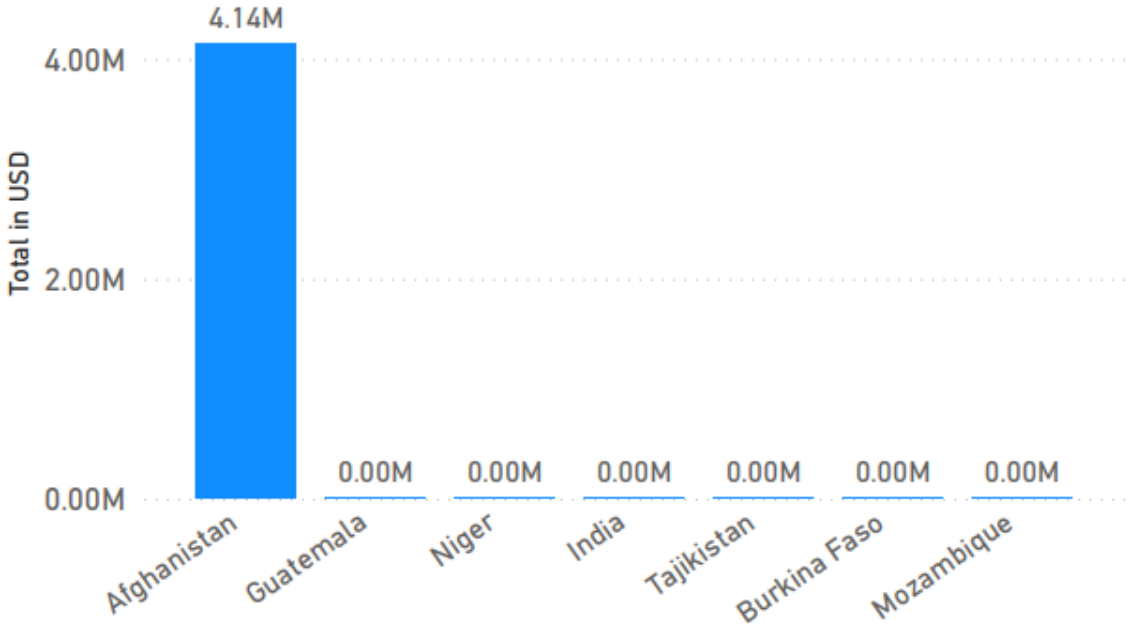
How we spent $7 million over 25 years (1992 → 2017)



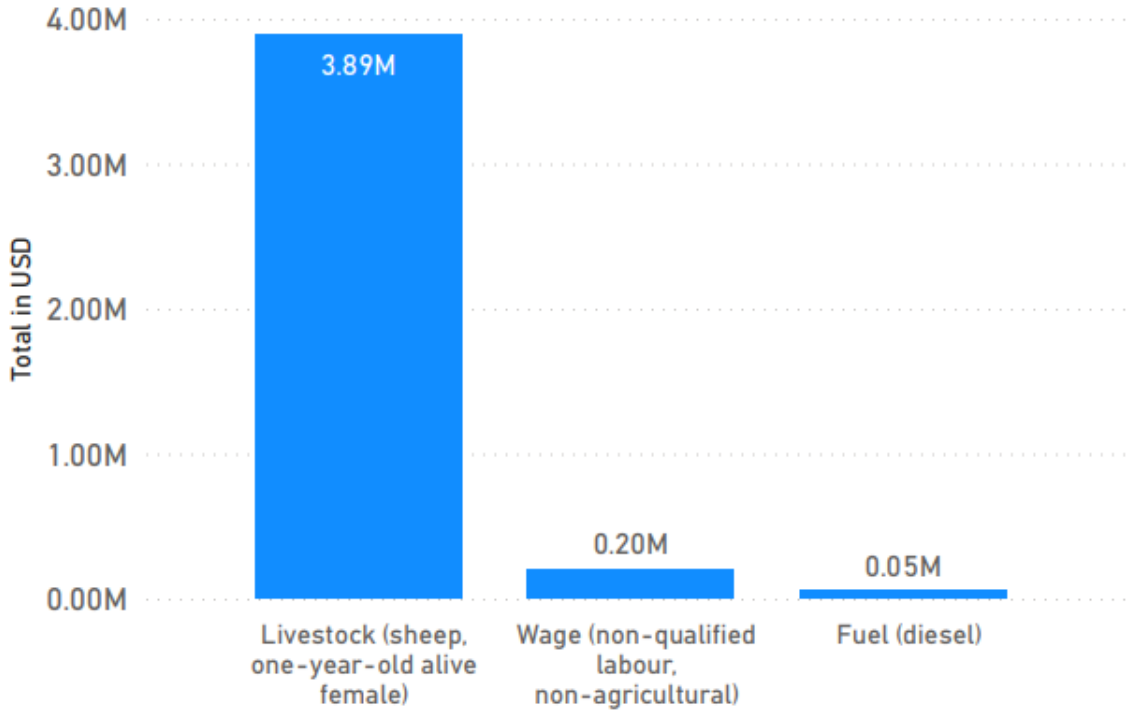
As we can see here, most of the expenditures were between 2000 and 2003. Let's see why?

We will check which countries have the expenditures in these years (2000,2001,2002,2003). Then will check the commodity too.

Set Up Report (Power BI) 02



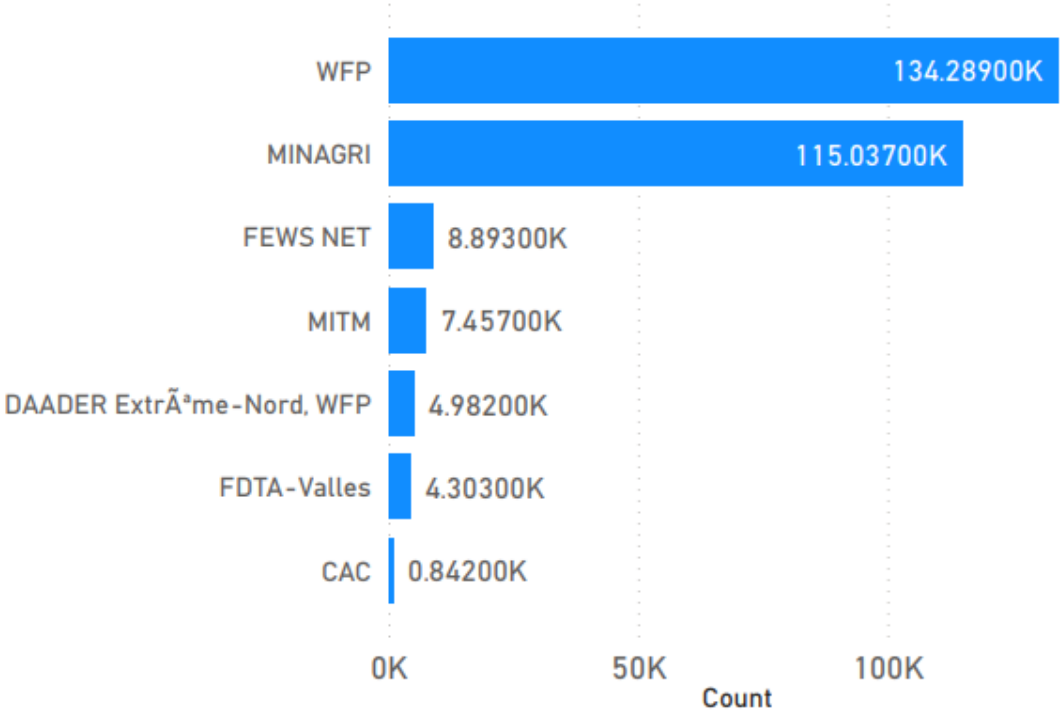
Afghanistan withdrew more than $4 million between 2000 and 2003.
At this time, Afghanistan was in a state of war, so it needed a lot of expenses.
So what kind of commodity they needed? Let's see



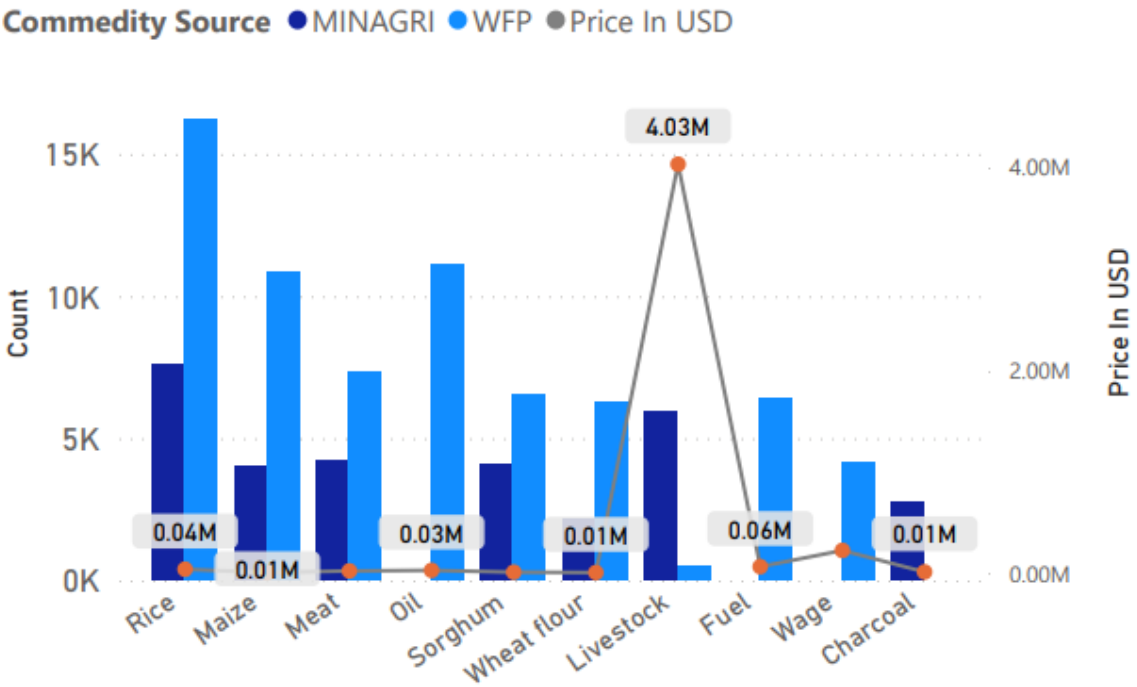
Livestock is withdrew more than $3.5 million between 2000 and 2003.

Set Up Report (Power BI) 03

Let's see the commodity source's distribution



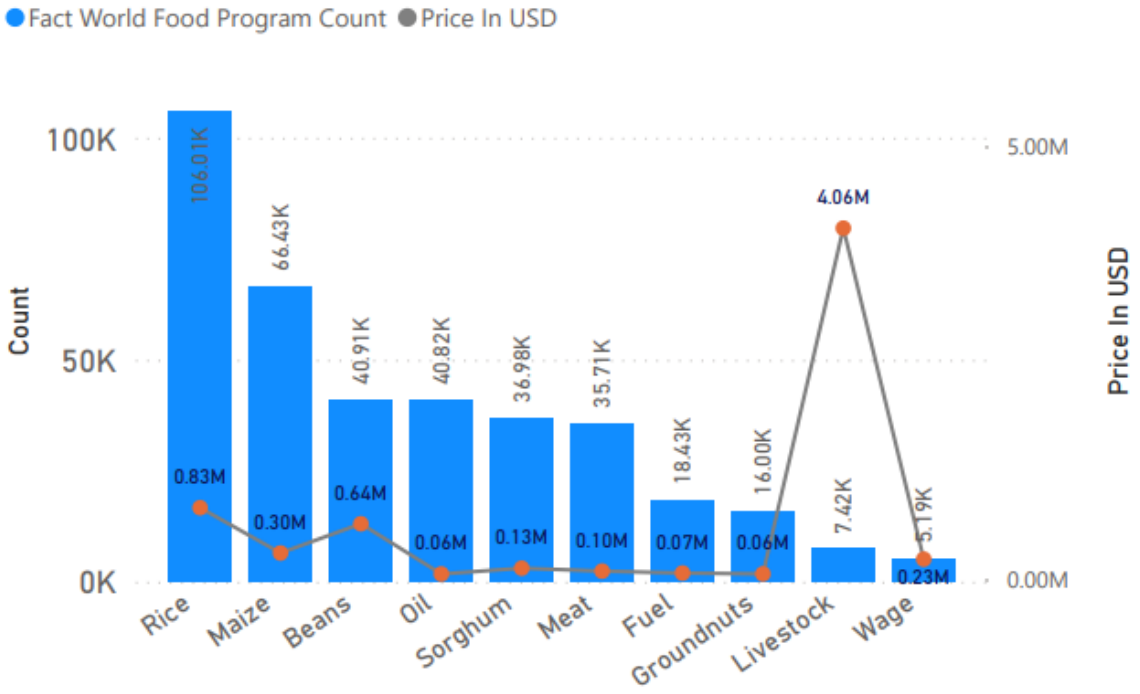
World Food Programme and **MINAGRI** are organizations that spent the most count in these years. (1992 ⟶ 2017). Let's see what kind of commodity they gave and how much they paid too.



Set Up Report (Power BI) 03

I think it is much clear now. Foods (Rice. Maize. Meat) are the most count and the Livestock are the most expenditures for **WFP & MINAGRI**

Let's see the commodity distribution and total expenditures for whole years



Summary

- Mot of the expenditures were for **Afghanistan** as a country and **Livestock** as a commodity.
- **WFP & MINAGRI** were the **biggest** commodity sources based on the count and expenditures.
- The period between 2000 to 2003 the spent is more than the half of the expenditures about (**$4.3 Million**).
- **Rice** is the highest count as a commodity

Techniques and Tools

01

Python

Used for data transformation, cleaning and getting third-party's API

02

SQL Management Server (SSMS)

Used to be the data warehouse for WFP's data and make DB creation

03

SQL Integration Services (SSIS)

Used to build ETL Pipeline, data transformation and move the data from table to another

04

SQL Analysis Services (SSAS)

Used to make a cube that has measures and dimensions tables + make a summary of the data

04

Microsoft Power BI

Used to make the visuals and analysis report to see the data as an interesting story

THANK YOU

Thank you for enjoying my first project in the field of data warehouse and business intelligence, and in the next projects you will be more excited, I am sure

CONTACT

Name: Abdelrahman Alarqan

Email: Abdelalarqan@gmail.com

Tel: (00972)-59-7824-195