World Health Organization

World Insurance Indicators

By: Dakota Brown

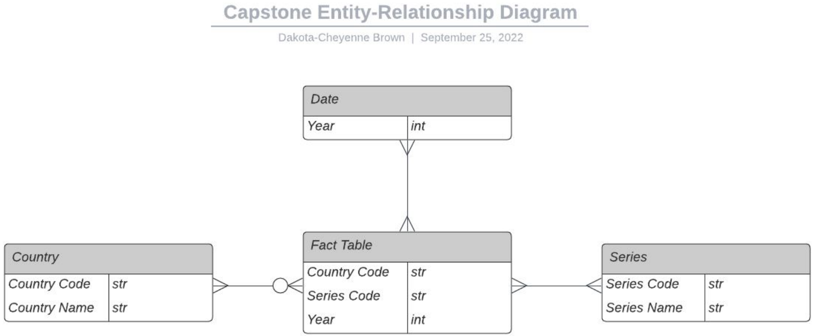
What is the problem to solve?

As someone working in the healthcare industry for the past two years, the biggest interactions I have had, other than with patients, is dealing with health insurances. Private health insurances rates have gone up with users still having to pay high copays and high out of pocket prices.

With health care costs rising and medical bankruptcies running rampant, how does the US healthcare costs compare countries with socialized healthcare?

Cleaning/Transformations

Before cleaning and transforming the data, I needed to be sure it adhered to a model. After Exploratory Data Analysis, I believed the data would work best in a star schema. From there I needed to create the auxiliary tables to the fact table as well as the fact table itself.



Capstone Entity-Relationship Diagram

Dakota-Cheyenne Brown | September 25, 2022

Date

Year	int

Country

Country Code	str
Country Name	str

Fact Table

Country Code	str
Series Code	str
Year	int

Series

Series Code	str
Series Name	str

Cleaning/Transformations cont.

I decided to use PySpark for the cleaning and transformations using AWS EMR. Using PySpark I was able to partition the data on different cores to tackle a bigger problem into smaller, easier to handle pieces.

I decided to use spot instances for both the cores and the masters to cut down on costs. When sizing the instances I chose a size big enough and computing power that could handle the data.

Testing

When testing my code, I wanted to assure that all the data I was extracting from the World Bank and World Health Organization were being pulled correctly.

The data needed to be:

- Taken from the World Bank’s API
- Extracted from zips and the correct CSV file chosen from three
- Loaded into an S3 bucket for storage before and after transformations

Through my testing, I was able to assure that the correct files were extracted and loaded into the S3 buckets before and after transformations took place

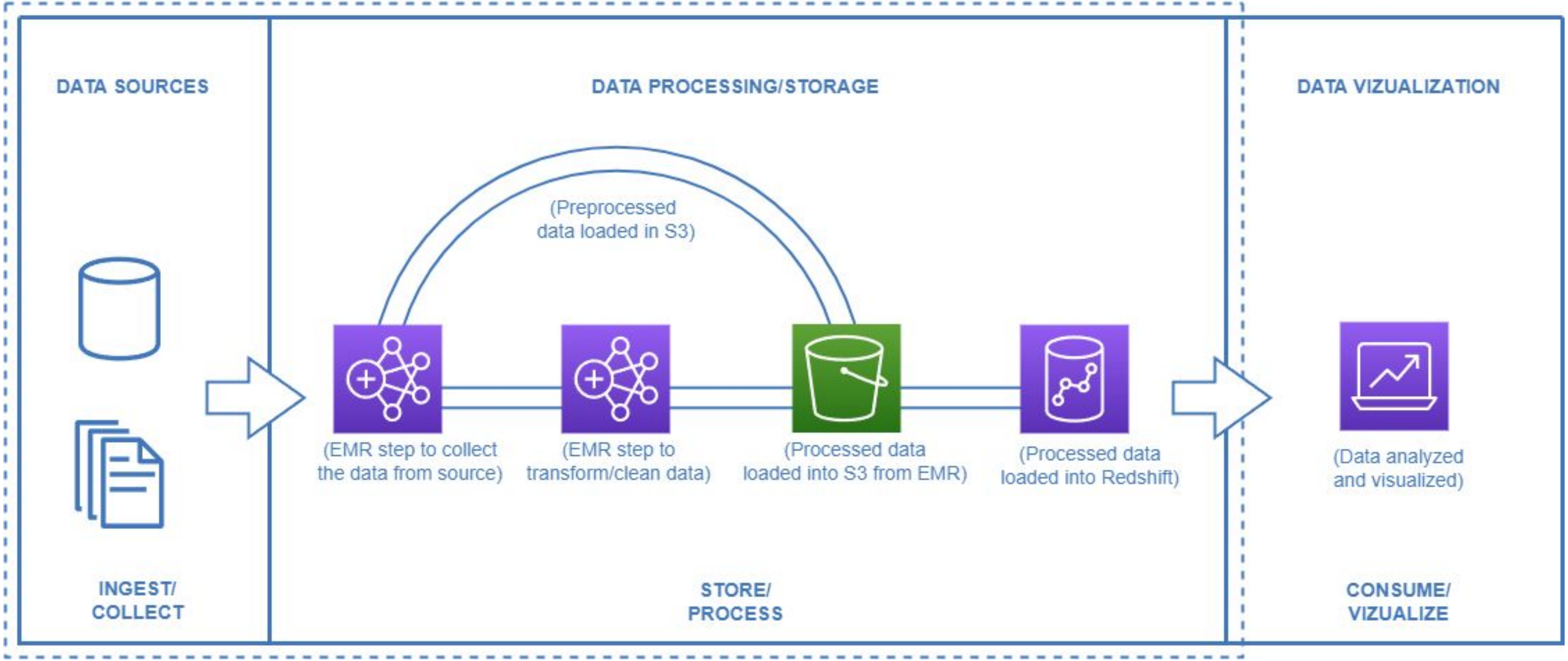
Architecture

When choosing the architecture for my project, I wanted it to be reusable, redeployable, and able to be ran by other engineers easily. I also needed to be sure the instances I used were correctly sized.

- I chose to use Apache Airflow for it use of DAGs and so it could be ran on a schedule
- From there an EMR instance was created that ran ETL code from S3
- The data was extracted into an S3, transformed in PySpark, and loaded back into S3
- All of this was containerized in Docker and ran on an EC2 instance

docker

Apache Airflow



DATA SOURCES

DATA PROCESSING/STORAGE

DATA VIZUALIZATION

(Preprocessed data loaded in S3)

(EMR step to collect the data from source)

(EMR step to transform/clean data)

(Processed data loaded into S3 from EMR)

(Processed data loaded into Redshift)

(Data analyzed and visualized)

INGEST/ COLLECT

STORE/ PROCESS

CONSUME/ VIZUALIZE

Metrics

When deciding on the metrics to visualize for this project, I had to make sure that I was getting the full picture that the data was presenting, and try to reduce my biases as much as possible.

Through that, I decided what I needed to track:

- Populations
- GDP
- Government spending on healthcare
- How much people spend out of pocket
- How much private insurances contribute to costs
- How the landscape has shifted over a 10 year window

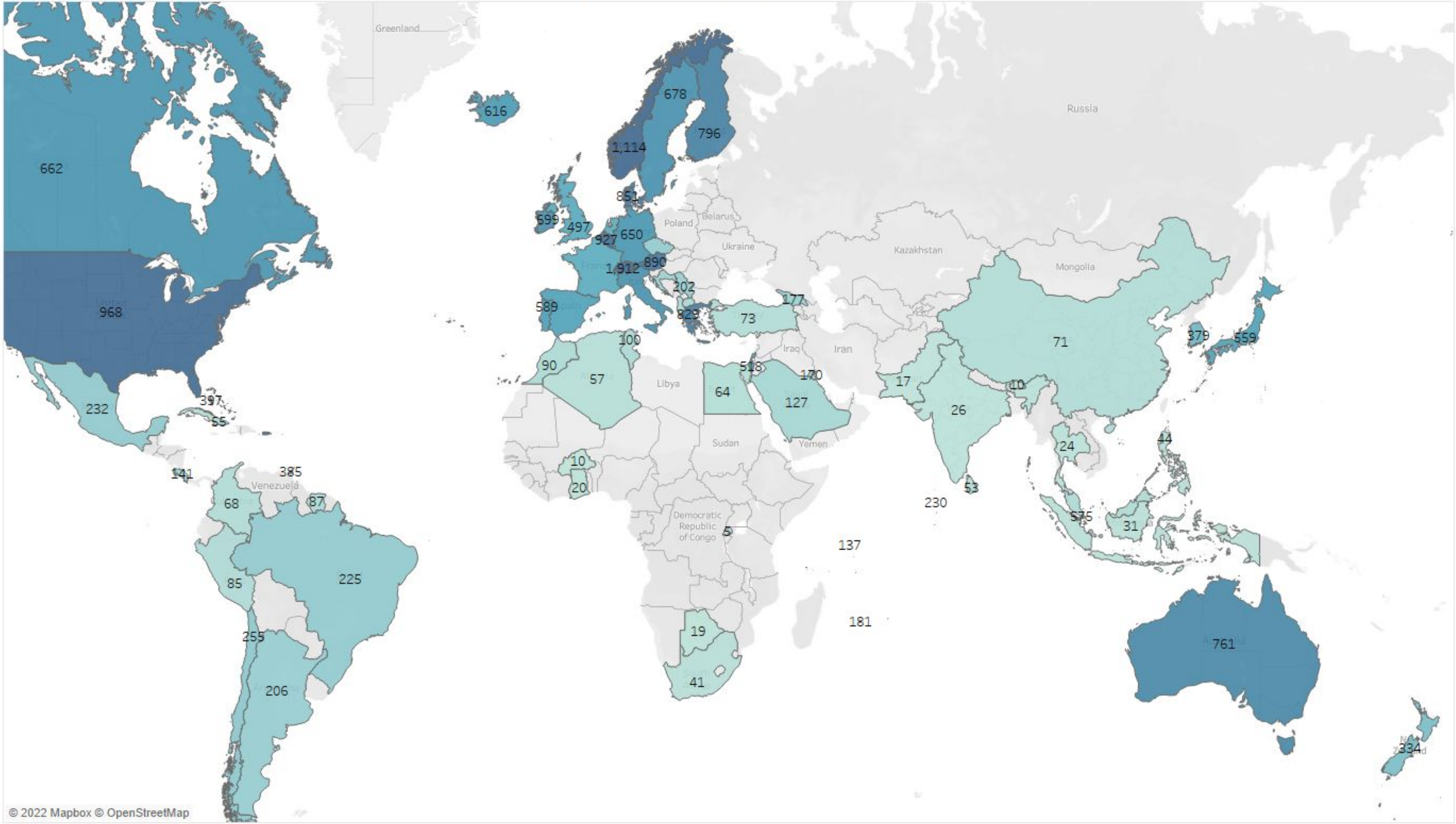
This link can be used to find all the metrics tracked for this project.

Insights

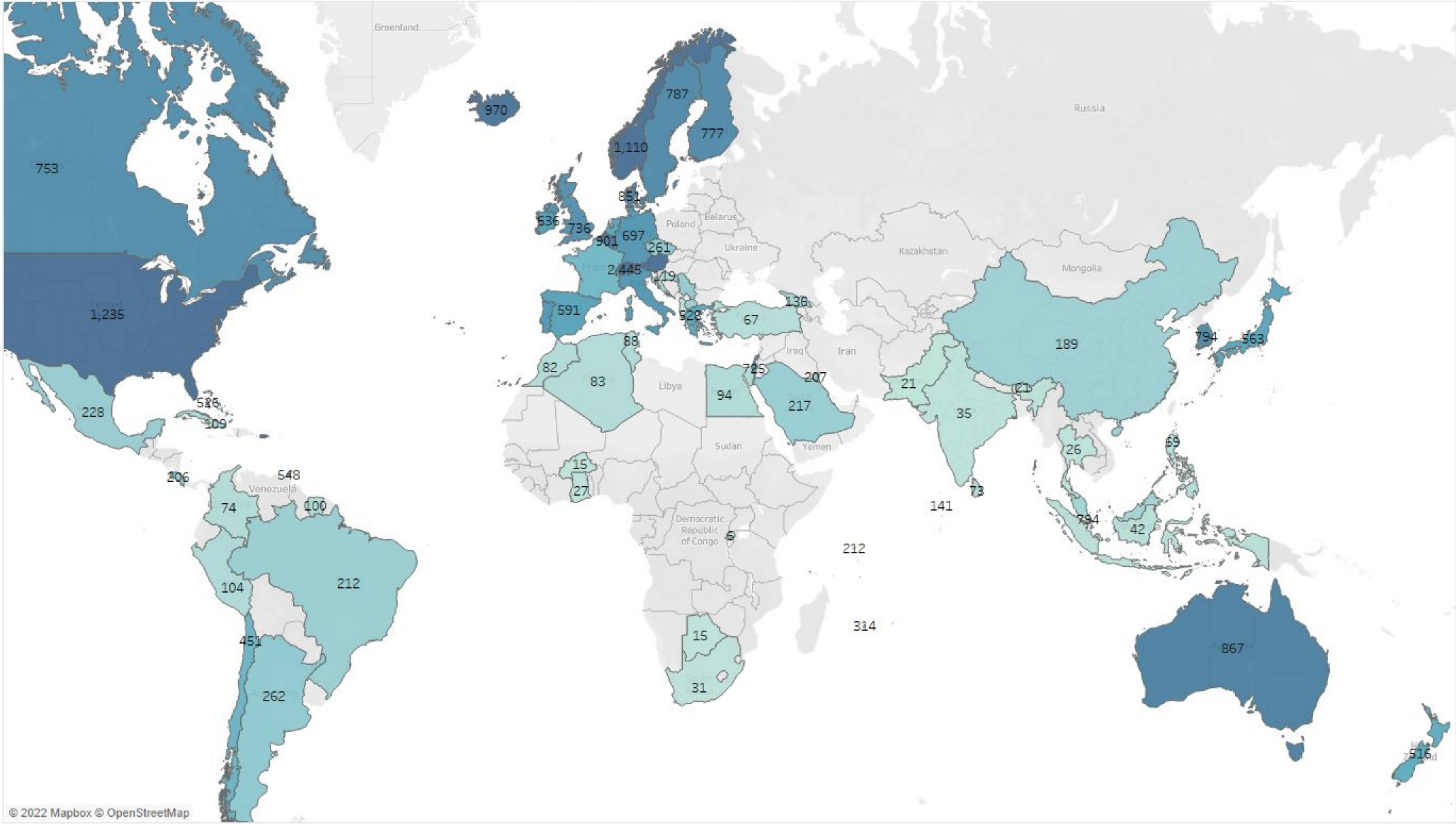
One of the insights I wanted to capture with this project was looking at the out of pocket expenditures for people in the US compared to those living in countries with socialized healthcare.

- Compared to countries with socialized healthcare the US went from third place to second in out of pocket costs for healthcare between 2009 to 2019.
- The other top three countries include behind Switzerland in the first place and Norway switching from second to third place between 2009 to 2019.
- During that time period,

2009 Out of Pocket Expenditure per Capita



2019 Out of Pocket Expenditure per Capita



© 2022 Mapbox © OpenStreetMap