



The Well Being of Women

By: **Betsy Kalkwarf**

Roopa Raghav,

Jessica Martin,

Paola Rivas



Reason Topic was Selected

As women, we care about women's well-being and strive to figure out what aspects of life most impact if women live long and happy lives.



Sources of data

3 Main Sources

1. LivWell (Kaggle)

2. Latitude/Longitude (Kaggle)

3. GDP (world bank data)

1. Livwell (Kaggle)

- a. LivWell is a global longitudinal database
- b. Provides a range of key indicators related to:
 - i. Women's socioeconomic status, health and well-being,
 - ii. Women's access to basic services, and demographic outcomes.
- c. <https://www.kaggle.com/datasets/konradb/wellbeing-of-women-in-52-countries?resource=download>

2. Latitude and Longitude (Kaggle)

- a. Latitude and Longitude for Every Country and State
- b. Provides the GPS coordinates for every world country and every USA state
- c. https://www.kaggle.com/datasets/paultimothymooney/latitude-and-longitude-for-every-country-and-state?select=world_country_and_usa_states_latitude_and_longitude_values.csv

3. GDP (World Bank Data)

- a. World Bank national accounts data, and OECD National Accounts data files
- b. <https://data.worldbank.org/indicator/NY.GDP.MKTP.CD>



Questions the team hopes to answer with data

- Is there a relationship between country demographics and aspects of life indicators (domestic violence rate, marriage age, years of education, and fertility rate) that impact women's overall well-being?
- Does GDP relate to these aspects of life?



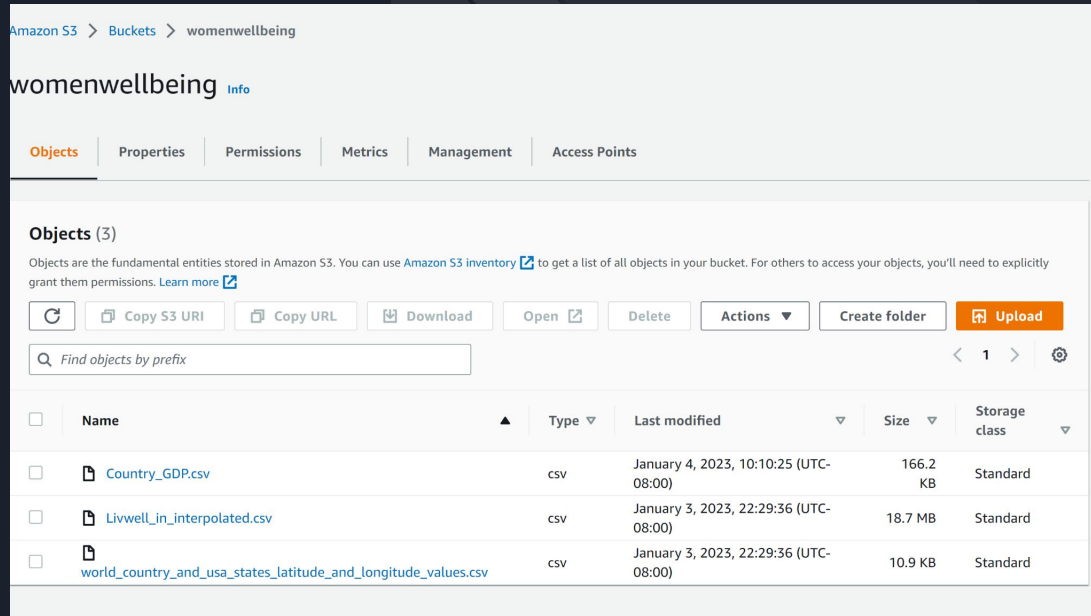
Data Exploration and Analysis



Data Exploration

Database Storage:

- Created a AWS database with PostgreSQL connection
- Created S3 buckets and uploaded the data sources
- Changed the visibility rules to public
- Created a PostgreSQL databases that connects to the AWS RDS



The screenshot displays the Amazon S3 console interface for a bucket named 'womenwellbeing'. The breadcrumb navigation at the top shows 'Amazon S3 > Buckets > womenwellbeing'. Below the bucket name, there are tabs for 'Objects', 'Properties', 'Permissions', 'Metrics', 'Management', and 'Access Points'. The 'Objects' tab is selected, showing a list of three CSV files: 'Country_GDP.csv', 'Livwell_in_interpolated.csv', and 'world_country_and_usa_states_latitude_and_longitude_values.csv'. Each file entry includes a checkbox, a document icon, the file name, its type (csv), last modified date, size, and storage class (Standard). Above the list, there are action buttons for 'Copy S3 URI', 'Copy URL', 'Download', 'Open', 'Delete', 'Actions', 'Create folder', and 'Upload'. A search bar is also present with the placeholder text 'Find objects by prefix'.

Amazon S3 > Buckets > womenwellbeing

womenwellbeing [Info](#)

Objects | Properties | Permissions | Metrics | Management | Access Points

Objects (3)

Objects are the fundamental entities stored in Amazon S3. You can use [Amazon S3 inventory](#) to get a list of all objects in your bucket. For others to access your objects, you'll need to explicitly grant them permissions. [Learn more](#)

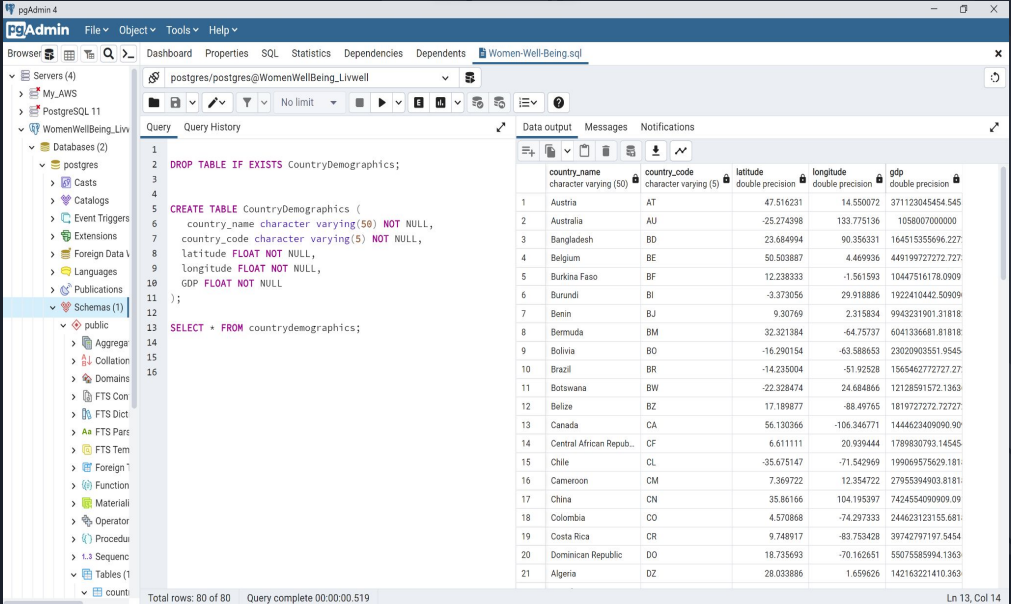
[Refresh](#) [Copy S3 URI](#) [Copy URL](#) [Download](#) [Open](#) [Delete](#) [Actions](#) [Create folder](#) [Upload](#)

<input type="checkbox"/>	Name	Type	Last modified	Size	Storage class
<input type="checkbox"/>	Country_GDP.csv	csv	January 4, 2023, 10:10:25 (UTC-08:00)	166.2 KB	Standard
<input type="checkbox"/>	Livwell_in_interpolated.csv	csv	January 3, 2023, 22:29:36 (UTC-08:00)	18.7 MB	Standard
<input type="checkbox"/>	world_country_and_usa_states_latitude_and_longitude_values.csv	csv	January 3, 2023, 22:29:36 (UTC-08:00)	10.9 KB	Standard

Data Exploration

Database Storage:

- Added latitude and longitude for maps visuals on the dashboard



The screenshot shows the pgAdmin 4 interface. The left sidebar displays the database structure, including 'Servers (4)', 'Databases (2)', 'Catalogs', 'Event Triggers', 'Extensions', 'Foreign Data Wrappers', 'Languages', 'Publications', 'Schemas (1)', and 'Tables (1)'. The main pane shows a SQL query being executed:

```
1 DROP TABLE IF EXISTS CountryDemographics;
2
3
4 CREATE TABLE CountryDemographics (
5   country_name character varying(50) NOT NULL,
6   country_code character varying(5) NOT NULL,
7   latitude FLOAT NOT NULL,
8   longitude FLOAT NOT NULL,
9   GDP FLOAT NOT NULL
10 );
11
12
13 SELECT * FROM countrydemographics;
```

The 'Data output' tab shows the results of the query, displaying a table with 5 columns: country_name, country_code, latitude, longitude, and gdp. The table contains 21 rows of data, including countries like Austria, Australia, Bangladesh, Belgium, Burkina Faso, Burundi, Benin, Bermuda, Bolivia, Brazil, Botswana, Belize, Canada, Central African Republic, Chile, Cameroon, China, Colombia, Costa Rica, Dominican Republic, and Algeria.

country_name	country_code	latitude	longitude	gdp
Austria	AT	47.516231	14.550072	371123045454.545
Australia	AU	-25.274398	133.775136	1058007000000
Bangladesh	BD	23.684994	90.356331	164515355696.227
Belgium	BE	50.503887	4.469936	449199727272.727
Burkina Faso	BF	12.238333	-1.561593	10447516178.0909
Burundi	BI	-3.373056	29.918886	192241042.50909
Benin	BJ	9.30769	2.315834	9943231901.31818
Bermuda	BM	32.321384	-64.75737	6041336681.81818
Bolivia	BO	-16.290154	-63.588653	23020903551.9545
Brazil	BR	-14.235004	-51.92528	1565462772727.27
Botswana	BW	-22.328474	24.684866	12128591572.1363
Belize	BZ	17.189877	-88.49765	1819727272.72727
Canada	CA	56.130366	-106.346771	1444623409090.90
Central African Repub.	CF	6.611111	20.939444	1789830793.14545
Chile	CL	-35.675147	-71.542969	199069575629.181
Cameroon	CM	7.369722	12.354722	27955394903.8181
China	CN	35.86166	104.195397	7424554090909.09
Colombia	CO	4.570868	-74.297333	244623123155.681
Costa Rica	CR	9.748917	-83.753428	39742797197.5454
Dominican Republic	DO	18.735693	-70.162651	58075585994.1363
Algeria	DZ	28.033886	1.659626	142163221410.363

Total rows: 80 of 80 Query complete 00:00:00.519 Ln 13, Col 14

Data Exploration

Database Storage:

- Rearranged columns
- Dropped null values into another dataframe
- Uploaded cleaned dataframe into database table in PostgreSQL for further analysis

	country_name	DV_phys_or_sex_partner_p	DM_age_mean	DM_age_marr_mean	ED_educ_years_mean	FF_TFR	PCA1	PCA3	PCA2	class	country_code	latitude	longitude	GDP
0	Armenia	0.539786	31.333770	20.173957	10.634626	1.777382	3.669187	0.091953	-2.105340	3	AM	40.069099	45.038189	8.706873e+09
1	Bangladesh	2.534889	25.935778	15.317056	3.770611	2.608617	-1.337056	-1.684766	-0.181221	1	BD	23.684994	90.356331	1.645154e+11
2	Benin	1.233981	28.747139	18.500306	2.392537	5.559406	-1.595005	-0.288428	-1.077806	1	BJ	9.307690	2.315834	9.943232e+09
3	Bolivia	0.000000	29.193333	19.904167	8.025083	3.904192	1.617289	-0.118519	-1.190013	3	BO	-16.290154	-63.588653	2.302090e+10
4	Burkina Faso	1.460385	28.893461	17.414231	1.301154	6.067308	-2.675948	-0.126313	-1.213747	1	BF	12.238333	-1.561593	1.044752e+10
5	Burundi	6.140357	27.701250	19.936250	4.436250	5.567125	-0.327420	-0.067633	-0.148522	1	BI	-3.373056	29.918886	1.922410e+09
6	Cambodia	16.259000	29.756852	19.504019	4.170741	3.502878	0.501395	0.365267	-0.536743	3	KH	12.565679	104.990963	1.362157e+10
7	Cameroon	43.352547	27.703113	18.511604	6.927547	4.782745	0.272630	2.468131	2.129353	2	CM	7.369722	12.354722	2.795539e+10

Data Analysis

Machine Learning:

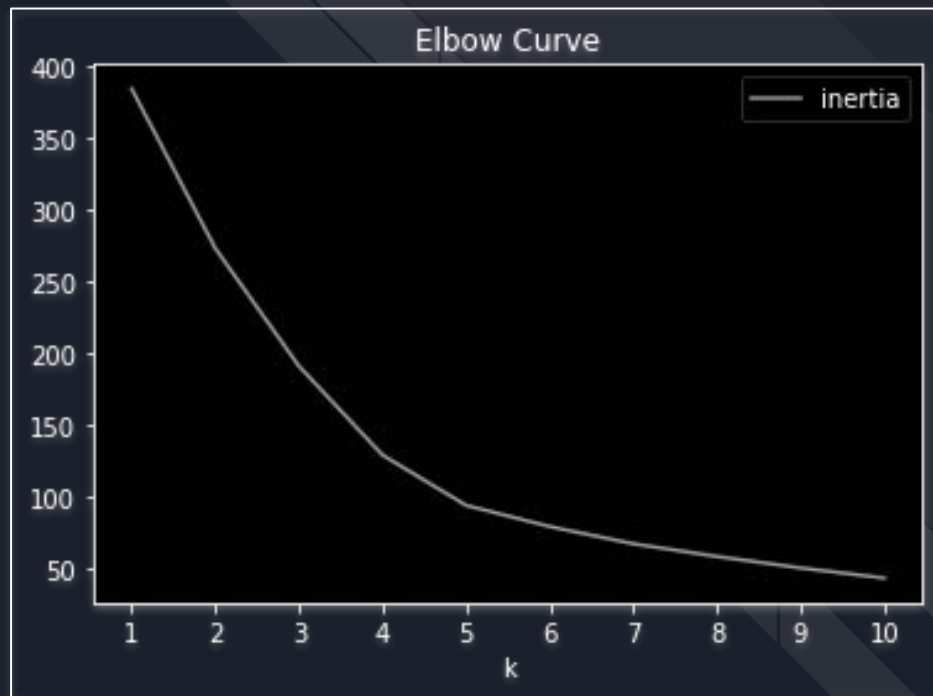
- Training and Test set up is Unsupervised
 - Chosen due to source not having any predictions
 - Wanted to cluster indicators chosen based on the country
- SciKit Learn is the machine learning library we'll be using to create a classifier



Data Analysis

Machine Learning:

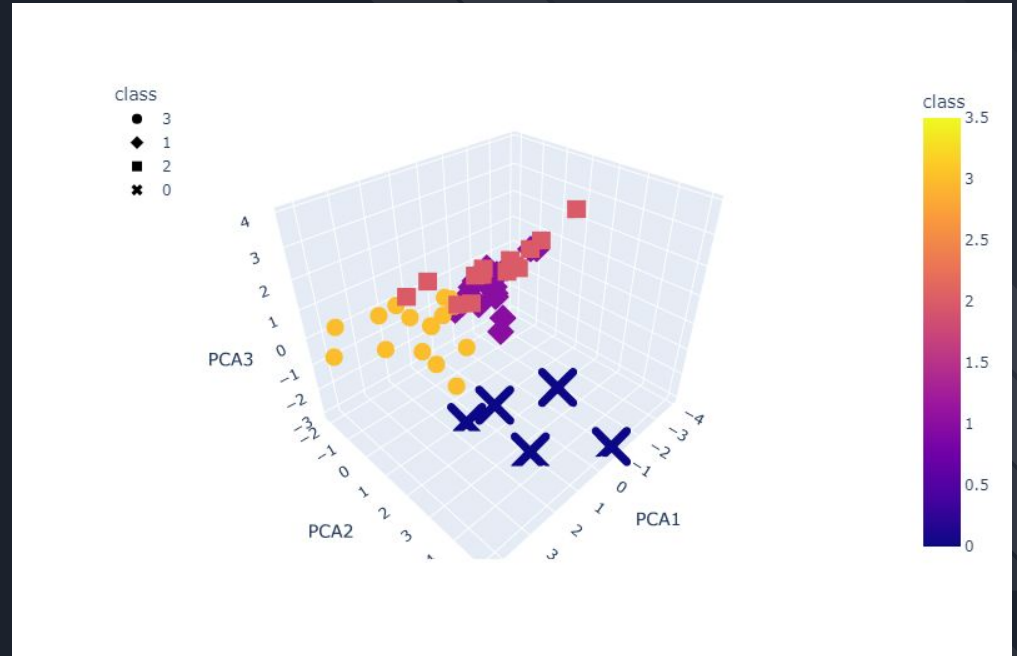
- Data was retrieved from database
- Set up ML model
- Scaled, fit and transformed the data
- Applied PCA for reduction
- Checked the Elbow Curve to decide the best K-value for clustering



Data Analysis

Machine Learning:

- 3D scatter plot created to check clusters



Data Analysis

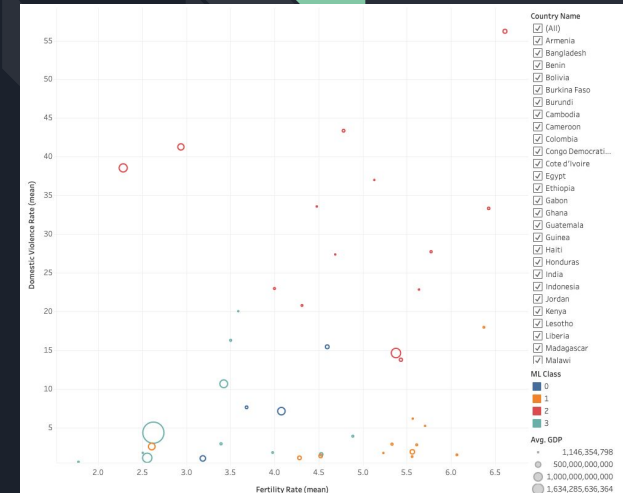
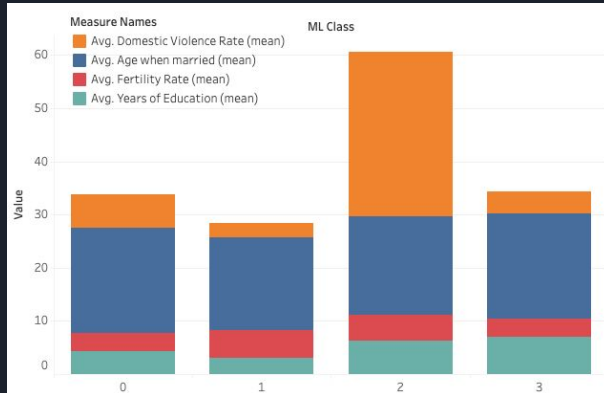
Dashboard:

The dashboard is hosted on [Tableau](#).

Domestic Violence



Fertility Rate





Thanks!

Are there any
questions?

