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?**

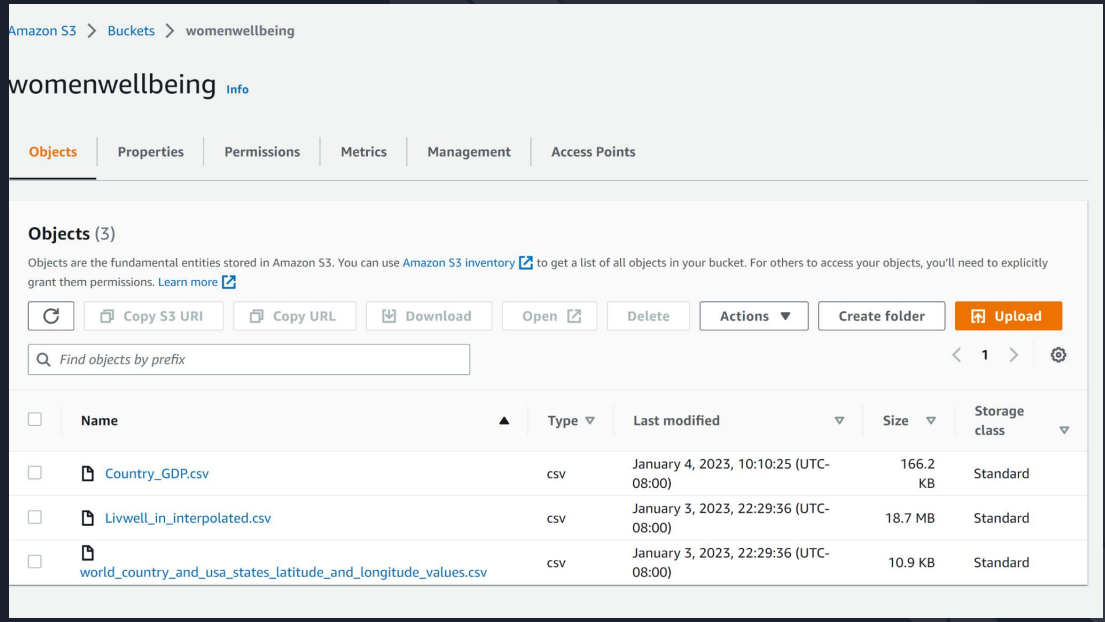
- **Does GDP relate to these aspects of life?**

Data Exploration and Analysis

Data Exploration

Database Storage:

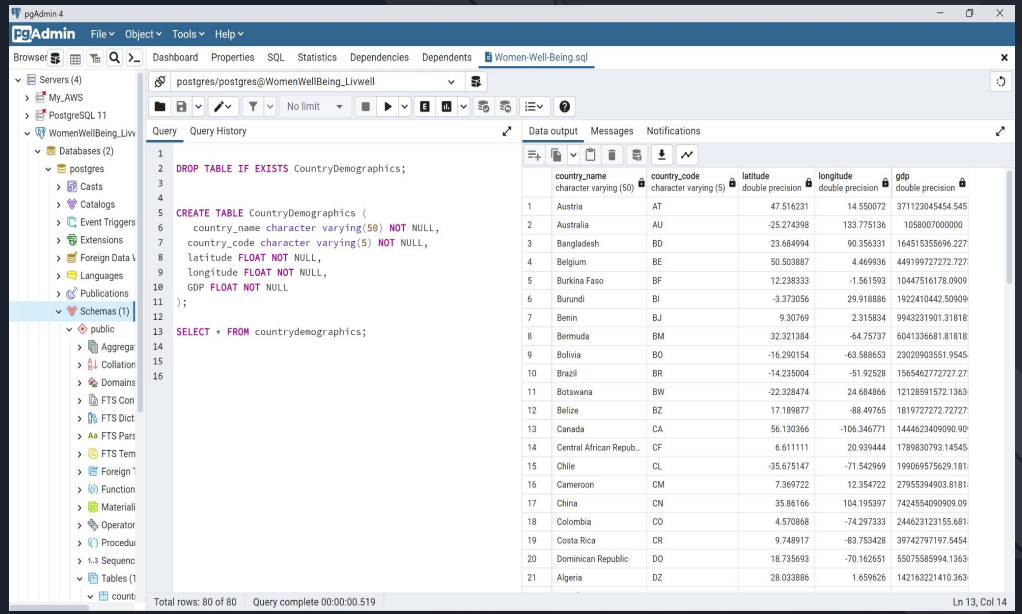
- Created a AWS database with PostgresSQL connection
- Created S3 buckets and uploaded the data sources
- Changed the visibility rules to public
- Created a PostgresSQL databases that connects to the AWS RDS



Data Exploration

Database Storage:

- Added latitude and longitude for maps visuals on the dashboard



Data Exploration

Database Storage:

- Rearranged columns
- Dropped null values into another dataframe
- Uploaded cleaned dataframe into database table in PostgresSQL 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

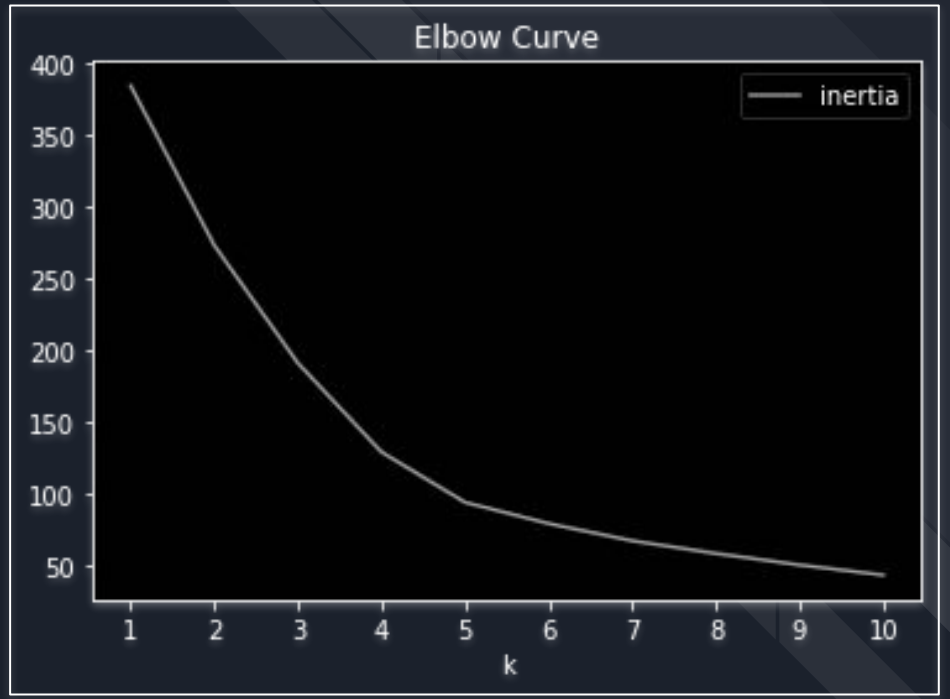
Machine Learning:

- 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:

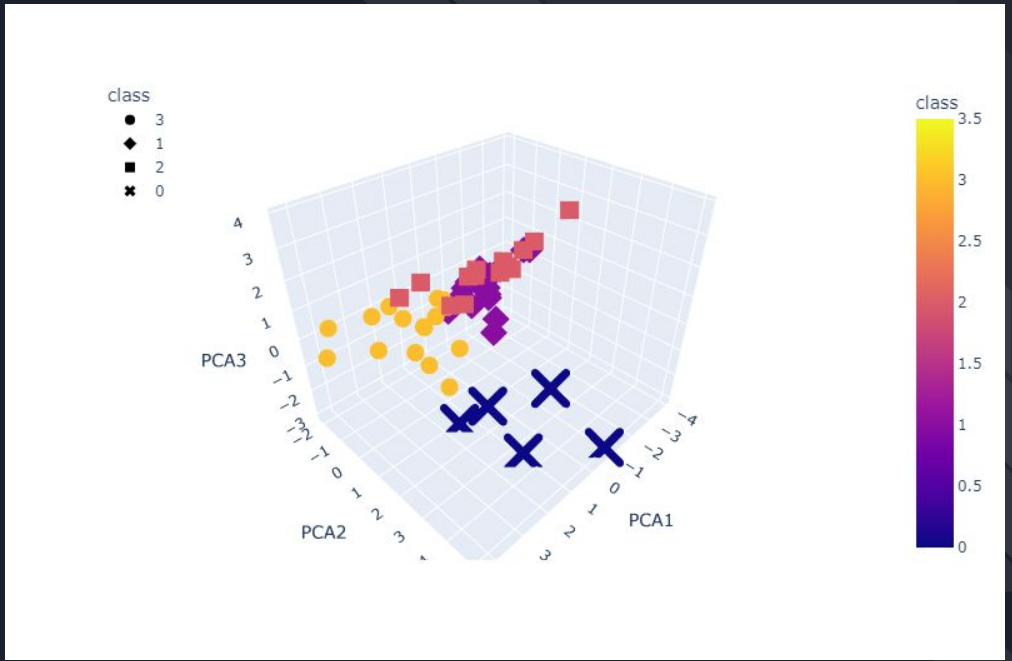
- Data was retrieved from database
- Set up ML model
- Scaled, fit and transformed the data
- Applied PCA for reduction
- Checked the Elbow Curve to decide the best K-value for clustering



Data Analysis

Machine Learning:

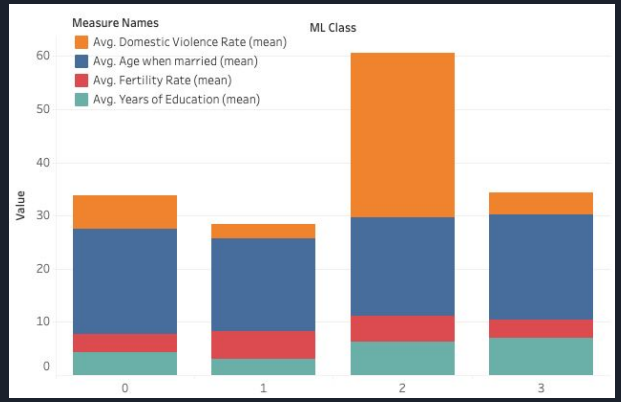
- 3D scatter plot created to check clusters

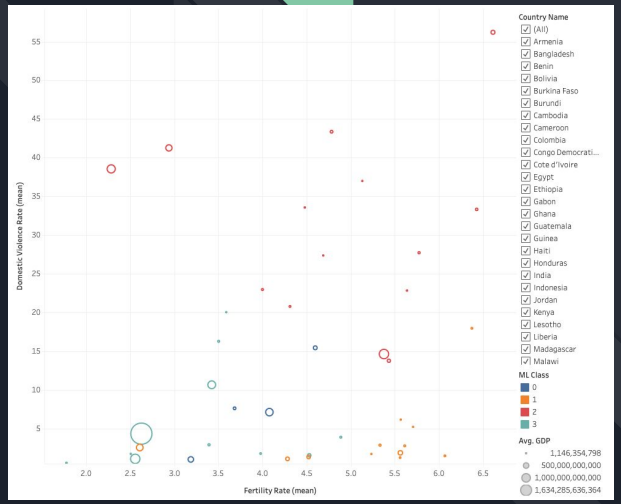


Data Analysis

Dashboard:

The dashboard is hosted on Tableau.





Thanks!

Are there any questions?