Life Expectancy Presentation

September 14th, 2022

7:00 - 7:10 pm

Rachel Ibach

Stephen Zimmermann

Scott Miller

Teddi Nutaitis-Lopez

Introduction

a. Meet the Team (1 minute): Slide 2

b. Introduction of the Topic (1 minute): Slide 3

Database

a. Data Processing and Database Creation (2 minutes): Slide 4 and 5

Machine Learning Model

a. Machine Learning Process: (2 minutes): Slide 6 and 7

Closing of the Topic

a. Tableau Visualizations and Dashboard (2 minutes): website map and data search functions; viz storyboard

b. Future Recommendations and Questions: (2 minutes)

Introduction

Meet the Team (1 minute): Slide 2

- Say everyone's names and to add us on linkedin so we can stay connected in the future

Introduction of the Topic (1 minute): Slide 3

- Throughout the centuries, life expectancy has increased due to improvements made to the environment in which a population lives in. This increase in life expectancy can be accounted for be analyzing factors that might contribute to a healthier life, such as economical factors, social factors, and mortality factors. The goal of our project was to be able to identify the effects of sanitation on worldwide life expectancy. We examined 193 countries within a decade long period.
- Both datasets come from the World Health Organization (WHO) and they cover the most recent health statistics of the world.
- Questions to Answer

a. Out of the countries with life expectancy changes, which factors affected this statistic the most?
b. Going forward, humanitarian efforts can be focused on what areas of the world?

Database

Data Processing and Database Creation (2 minutes): Slide 4 and 5

- Main Dataset: Post Cleaning

- 193 Countries, 2005-2015

- 2023 Data Points with 22 Features (minus target feature)

  ex. GDP, Alcohol Consumption, Population, Developing/Developed

- 7 separate datasets

  → 5 Added water and sanitation features

  → 2 were to clean/replace features

- Issues:
- Nulls, Zeros, String to Ints, and Inconsistencies
- Population changed by factor of 10x year to year
- Getting new clean data for 'population' and 'GDP' required conversions for wide to long (format)

Primary KEYS: Year and Country

- Country names were inconsistent

- Was a iterative process with the EDA and ML

Machine Learning Model

Machine Learning Process: (2 minutes): Slide 6 and 7

- Feature selection was done by considering P-value and VIF. Goal was to get p-value below 0.05 for all features and vif below 5.0
- Removed 1 feature at a time and updated model, by the 8th model, reached the threshold and had our list of features to use

- Validation with test set came back with accuracy/r-squared value of 0.707 (total of 8 features)
- Introduced the sanitation files to the original dataset to see if stronger results. Followed same process and this time to get VIF below 5 and p-value below 0.05, this time feature selection was trimmed down to 4 features and but accuracy improved by about 10%. The validation r-squared was .802

Random Forest Regression

- Dropped country and used getdummies function on status and then used all columns for the model
- Ended up with a RMSE(root mean square error) of 2.014 and an accuracy score of 99.07 (possible overfitting) and the validation or testing had a 98.12% accuracy.
- Density Plot to show actual life expectancy versus the predicted life expectancy of the testing set

Closing of the Topic

- Tableau Visualizations and Dashboard (2 minutes): website map and data search functions; viz storyboard
- Tableau visualizations: shows that as the sanitation factors become more implemented in countries, the life expectancy increases; many interesting visualizations came about from the data
- We have an interactive map that shows each countries name and their life expectancy along with the sanitation statistics listed below.
- We have an option to search through the countries with multiple filters that can altered

Future Recommendations and Questions: (2 minutes)

- Our results were that an increase in sanitation factors such as: hand washing facilities in the home, increase in basic sanitation features, drinking water services, and safe sanitation services, would lead to an increase in life expectancies in both developed and developing nations (but it was was most apparent in developing nations)
- There were a lot of interesting results from this analysis and we believe it could be useful for humanitarian efforts: focusing on the factors that most affect the life expectancy
- If we were going to do it again, we would have found a more complete dataset and potentially analyzed other features that affect life expectancy (such as diseases, income, and schooling)
- Thanks for listening, any questions?