CMSE 830 Mid-Term Project: Peer Reports

Angelica Louise Gacis

Date: October 26, 2022

Presenter 1: Shuangyu Zhao

Goal of the project: The goal of her project is to verify the diagnosis of heart disease presented in the data set. She used EDA to explore the features related to heart disease and also applied a classification model.

Learnings from the app: I learned about the different factors that are attributed to heart disease as listed in the app's introductory page. In the second page, which is the EDA part, I learned about the attributes that are highly correlated with diagnosis such as "exercise induced angina" and "ST depression induced by exercise relative to rest". I also learned how she identified the dataset as a good fit for creating a classification model. In the third and last part of the app, I learned about the KNN classifier model which is what Ms. Zhao used for predicting diagnosis based on given attribute values.

Visualizations' aid to narrative understanding: Most of the visualizations helped me in understanding the story. At first, I was confused with navigating the 3D scatter plots, but I understood the purpose of it when I saw that one of the axes was categorical. If those were in 2D, they would only appear like lines.

Visualization strategies used: One of the strategies she used was to add color or highlights to the data frame of expected values. I thought that it was really helpful and easy to understand. The colors for the plots were also consistent, hence not distracting. She also used some tooltips for the interactive plots which were helpful especially for the scatter plots.

Plot libraries used: Ms. Zhao used seaborn plots for the heatmaps and boxplots for outliers, altair for the 2D scatter plot, plotly for the 3D scatter plots, and streamlit for the bar chart of classification. Overall, the plots used were appropriate, but it can be improved with consistency in the visualization type or theme, at least from my end.

Preprocessing: She included her preprocessing in the web app by showing us the heat map indicating whether there are missing values or not. Personally, I think that the heatmap is only necessary to show patterns for a large number of missing data. Because her data set did not have any missing values, the heat map became an additional visualization. On the contrary, there were outliers shown in the box plots and I agree with the removal of those which is what she did in her preprocess.

App Design: For the app design and features, she used a navigator on the sidebar for the different pages of the app. There was also a dropdown for the stacked bar chart. I think that the design choices were appropriate and well decided upon especially for the prediction page which is very straightforward and clean. For improvement, the captions in the data analysis part may be shortened a bit.

Date: October 26, 2022

Presenter 2: Pavan Yachamaneni

Goal of the project: The goal of his project is to use EDA in exploring the different countries' GHG emissions from 11 sectors from 1990 to 2019. The web app aims to answer which sectors from different countries contribute the most to pollution through their GHG emissions. Mr. Yachamaneni also used the % of growth in each country from 1990 to 2019 to help in explaining outlying patterns that might affect GHG emission.

Learnings from the app: I learned that there are around 20 countries that contribute to approximately 80% of the GHG emissions worldwide. Most of the sectors that contributed were buildings, electricity, manufacturing, and transportation. The general trends are fluctuating because of different events that impact the world.

Visualizations' aid to narrative understanding: I am impressed with the visualizations used because they were all simple and straight to the point.

Visualization strategies used: I especially liked the horizontal bar graphs wherein bars with higher values were shaded darker. This made the plots really crisp, clean, and easy to understand. There were also instructions at the top which were not long, but were helpful.

Plot libraries used: Mr. Yachamaneni used altair for all of his plots. I feel like they were good choices. Personally, I find altair plots a bit distracting because of all that is going on, but this app showed me that a clean altair plot actually presents beautiful visualizations. If possible, the plots can be made cleaner such as removing the grid lines at the back, removing the markers, and applying value labels instead for those that are wished to be highlighted.

Preprocessing: Although not included in the web app, he applied outlier removal and transformations to the data set for the app's usage.

App Design: Overall, I think that the app design is really good. I think that the intentional choice not to use a sidebar helped in a way that everything was already there in front of the user, so less clicking has to be done. Mr. Yachamaneni also did no

Date: October 26, 2022

Presenter 3: Souradip Biswas

Goal of the project: This project uses EDA to explore the different features attributed to heart failure and heart conditions such as high blood.

Learnings from the app: The EDA mostly explores differences of observations between males and females as well as among different age groups. Heart disease has the chance of happening equally between males and females. However, males tend to have higher cholesterol levels. For age, the older people get, the lower their heart rate is. I also learned that people with fasting blood sugar of more than 120 mg/dl are significantly more at risk for heart disease.

Visualization: There was a technical difficulty deploying the app before the deadline of this report, so the link was not available after the presentation. However, from what I can recall, the visual aids were straightforward, although some of them can be improved such as the scatter plot for a binary variable. Other types of plots can be explored and extra ink can be removed further.

Plot libraries used: Mr. Biswas used altair for all his plots in the app.

Preprocessing: The data set was from Kaggle and there was no missing data.

App Design: Overall, I thought that the app design was sufficient to convey the story that Mr. Biswas wanted to tell

Date: October 28, 2022

Presenter 4: Sangeetha Suresh

Goal of the project: This project uses EDA to explore the relationship between the features of baseball games and game wins over the last ten years. According to Ms. Suresh, this app aims to help in strategizing the playing style of baseball players.

Learnings from the app: I learned that the divisional winners generally have equal numbers of defensive and offensive players. Furthermore, world series winners are mostly defensive teams, but those who have higher hitting metrics are those on the offensive teams. This learning suggests that focusing on stopping runs rather than the scores can actually do better. I also learned that there is a strong relationship between wins and runs.

Visualizations' aid to narrative understanding: There are several plots used in the app and I think that the choices are appropriate and helpful. However, the strategies on how the plots were arranged could be better. For example, the bar plots for the "Division winners and their play style" in the Play Style section can be collapsed wherein the user can choose among the three plots to show. This would lead to a consistent number of one or two plots per section.

Visualization strategies used: I really like the uses of orange and gray in the Wild Card section. The same strategy could have been applied to the other sections as well. Besides the Hi-Plot, none of the other plots were interactive which actually makes the app simpler, straightforward, and easy to understand.

Plot libraries used: The libraries used were Matplotlib and Altair, and I think they were good choices as well.

Preprocessing: Ms. Suresh explained that even though the data did not have missing values, the hitting metrics had to be determined explicitly as part of the preprocess.

App Design: Overall, I think that the app was clean and straightforward. The observations were brief and direct as well. The sidebar was utilized well for its purpose. Besides the ones I have mentioned above, the existing dropdowns were placed properly.