You've passed your capstone project!

SUMMARY

The impact of this project is significant and it can contribute to explore if predictions in certain illnesses can be made related to habits, nutrition, BMI and blood work using statistical analysis, machine learning and data science. This project uses National Health and Nutrition Survey (NHANES) dataset that assesses the health and nutritional status of people in the United States and show the applications of Data Science in forecasting and future prediction. Experiments are designed and executed in order to generate actionable insights. You have tried all the standard machine learning algorithms which is a good practice to learn and evaluate them. Experiments have shown good prediction results using Logistic Regression algorithm which can predict whether certain data points and information can lead to Asthma with an accuracy of 60%. Second part of project utilizes K-Means clustering technique to segment different groups of people by finding the unique pattern in their attributes. From the application point of view, these segments can be used to provide personalized treatments to each group. Overall you did a great job completing this project.

COMPLETION

All the deliverable are available on GitHub including Initial Project Ideas, Project Proposal, Data Wrangling, Exploratory Data Analysis, Data Story, Milestone Report and In-depth Analysis on Python Notebook. A problem was selected that has practical application, the value for a client, real or imaginary is justified and how the outcomes can be used are described, and the project was scoped appropriately for the course.

PRESENTATION

The report and presentation are clear and comprehensive. The submission includes code for the project. The submission includes a slide deck presentation. The submission includes a report based on summary reports from all previous project submissions. The github submission is complete, each part is organized in a separate repository. All intermediate capstone projects (Initial Project Ideas, Project Proposal, Data Wrangling, Exploratory Data Analysis, Data Story, Milestone Report, In-depth Analysis) are submitted.

UNDERSTANDING AND PROGRESS

You selected a problem which has practical application in understanding risk factor for certain illnesses that are related to habits, nutrition, BMI and blood work. The data utilized in the project shows understanding of how to acquire, merge, wrangle and clean data. National Health and Nutrition Survey (NHANES) is a well-chosen organization for data source and relevant to the problem. The technical framework you designed covers all the main components of Data Analytics such as (1) Select and utilize appropriate algorithms for this application (2) Apply standard EDA techniques such as histograms, scatter plots and hypothesis testing (3) Utilized machine learning algorithms for prediction and segmentation (4) and confusion matrix classification accuracy as evaluation metric/technique for your algorithm. Your project demonstrates skill in communication of the project results that includes (1) Utilizing data visualization (2) Developing a story to present your hypotheses (3) and outcome analysis in a well articulated data story relevant to the target audience. Your code is well documented, clean and contains output for each step. Each step has its own code file which is good for code usability. The report is clear and comprehensive: stating the problem well, describing the data acquisition and wrangling challenges, articulating how EDA was implemented, and providing clear outcomes, recommendations, and next steps. The slide deck presents the motivations and recommendations clearly and impactfully, utilizing visuals. It includes 1-3 clear recommendations well-justified by the analysis.