Week 1 summary

Implementation

Day 1: Data Exploration and Image Processing

On the inaugural day, the team consisting of a Data Analyzer, Data Provider, Documentor, and Note Taker embarked on their journey. They began by exploring image data representation and processing, gaining insights into the storage and processing of digital images. The outcomes of the day included a foundational understanding of image data representation and basic image processing techniques.

Day 2: Excel Data Analysis and Visualization

Day 2 saw the team diving into data analysis using Microsoft Excel. They worked with a small Excel dataset, performing calculations to derive descriptive statistics and created a bar chart for data visualization. This marked the transition from image data exploration to practical data analysis and visualization skills in Excel.

Day 3: Introduction to DataFrames in Python

On the third day, the team learned about DataFrames in Python and their broad applications in data analysis. They conducted fundamental operations on DataFrames, including data selection and manipulation, setting the stage for more advanced data analysis using this versatile data structure.

Day 4: Data Analysis and Model Preparation

Day 4 focused on data analysis and model preparation for the Fake Bill Detection project. The team segmented the dataset into dependent and independent variables, created dummy variables, and removed specific attributes. The data was split into training and testing sets, laying the groundwork for developing a logistic regression model.

Day 5: Model Validation and Assessment

The fifth day was dedicated to validating the model constructed on Day 4. The team assessed the model's performance through essential tools such as the confusion matrix, ROC curve analysis, and the computation of KS statistics. This ensured the model's effectiveness for the Fake Bill Detection project.

Outcomes of Week 1

Throughout Week 1, the team acquired a diverse set of skills and knowledge. They transitioned from image data exploration to practical data analysis and visualization in Excel. Subsequently, they delved into the world of Python DataFrames, which opened up opportunities for more in-depth data operations. They also embarked on model development and validation, critical for the Fake Bill Detection project. These outcomes provide a strong foundation for advanced data analysis tasks in the weeks ahead, demonstrating the team's growing expertise in data analysis techniques. The journey is off to a promising start, with productivity and insights on the horizon.