**Week 2**

**Summary**

**Milestones achieved**

* Examined data through visualization and analysis techniques.
* Generated statistical summary
* Feature correlation to output class(es).
* Generated Scatter plot
* Dealing with Categorical data
* Fixed problems like missing values, errors or outliers.
* Apply pre-processing or normalization procedures.

**Conclusions**

**For Data set 1**

* Features like Online Boarding, In-flight Entertainment, and Seat Comfort are important factors in customer satisfaction and should be focused on for improvements.
* Negative Influencers: Being in Economy Class or on Personal Travel has a strong negative impact, meaning that premium or business services are more likely to lead to satisfaction.
* Neutral Features: Some features like Gender\_Male (0.011236), Gate Location (-0.002793), or Departure Delay (-0.050740) have very low correlation, meaning they do not significantly affect satisfaction.
* Hence**,** based on the correlation values, we can decide which features to prioritize in our model. Features with higher positive correlations could be key predictors, while features with low or negative correlations may be less useful.
* Preprocessing was applied to deal with missing data as they can negatively affect model performance
* Different features in our dataset are on different scales (e.g., Age ranges from 18-60, while Flight Distance ranges from 0-5000+). Some machine learning algorithms (e.g., KNN, SVM, neural networks) are sensitive to the magnitude of values and will give more weight to features with larger numerical ranges.
* Hence, we normalize or standardize the numerical features so they are on the same scale.

**For Data set 2**

* We have observed the same correlation behaviors as the data set 1. Hence concluded for **Improving Customer Experience**, these insights can inform business strategies. For example, improving online boarding processes or in-flight entertainment might increase customer satisfaction, while addressing issues in economy class services could reduce dissatisfaction.

**References**

* <https://www.kaggle.com/datasets/teejmahal20/airline-passenger-satisfaction/data>
* <https://ebookcentral.proquest.com/lib/hw/reader.action?docID=30168989&ppg=67>
* Géron, A 2022, Hands-On Machine Learning with Scikit-Learn, Keras, and TensorFlow, O'Reilly Media, Incorporated, Sebastopol. Available from: ProQuest Ebook Central. [29 September 2024].
* <https://github.com/ageron/handsonml2/blob/master/02_end_to_end_machine_learning_project.ipynb>

**Challenges**

* Struggling with concepts and understanding of data sets
* Understanding workflow

**Next Steps**

* Examining features by training classifiers and applying evaluation metrics
* Start studying the scikit-learn library of Python.
* To Search more about machine learning models to get some clarity about our upcoming task.