Week 10 Deliverables

Submit a pdf document and EDA ipynb file which should contain following details:

Team member's details : Data Collector, Katrina Kirby, katkirby9217@gmail.com, United States, Virginia Commonwealth University, Specialization (Data Analyst)

Problem description: XYZ company is collecting the data customer using google forms/survey monkey and they have floated in number of forms on the web. Company wants to create a pipeline which will collect all the data of these google forms/survey monkey and visualize the data in the dashboard. Company wants clean data and if there is any data issue present in the data then it should be treated by this pipeline (duplicate data or junk data). dedup check should be performed on the email id of the customer

GitHub Repo link: https://github.com/kirbykt/Data-Analyst-Project

EDA performed on the data:

- **.isnull().sum() analysis**
- **Boxplots for outlier checks**
- **Countplots of categorical columns**
- **Grouped bar plots**
- **Pie chart for issue resolution**
- **Summary of NA values and strategy taken**
- **Issue resolution rates**
- **Locations of highest service ratings**
- **Timestamp addition for audit purposes**
- **Recommendations by satisfaction levels**
- **Handling missing values, data, and duplicates**
- **Distribution of ratings and satisfaction**

Final Recommendation:

- Most customers report satisfaction levels of 4 or 5 and are likely to recommend the product
- All service-related issues were resolved, with high support experience scores (especially in Midlothian, VA)
- No major outliers were observed in satisfaction or support scores, and missing values were minimal.

  It is recommended that:

- Continuing to emphasize service quality, particularly in areas showing high support ratings
- Encouraging product reviews and referrals through loyalty incentives
- Monitoring satisfaction trends over time and across locations to identify emerging patterns