**Week 1**

**Summary**

**Milestones achieved**

* Group Formation
* Python set up; GitHub set up
* Explored 3 data sets from at least two sources. Discussed among the groups, compare similarities and differences.

**Conclusion**

* Concluded Instances, Attributes—Ratio, Interval, ordinal, Nominal and Classes for the observed data sets.

**For data set 1 (Hurricane data)**

* Nominal: ID, Name, Status, Event.
* Interval: Date
* Ratio (Continuous): Time, Latitude, Longitude, Maximum Wind, Minimum Pressure, Wind attributes (Low, Moderate, High for different directions)

**For data set 2 (airline data)**

* Nominal: ID, Gender, Customer Type, Type of Travel, Class.
* Ratio (Continuous): Age, Flight Distance.
* Interval (no true Zero): Departure Delay, Arrival Delay.
* Ordinal: Departure and Arrival Time Convenience, Ease of Online Booking, Check-in Service, Online Boarding, Gate Location, On-board Service, Seat Comfort, Leg Room Service, Cleanliness, Food and Drink, In-flight Service, In-flight Wifi Service, In-flight Entertainment, Baggage Handling, Satisfaction.

For dataset 3: (Image Dataset)

* The image dataset contains two folders as train & validate
* Each dataset contains 7 classes: Angry, Happy, sad, fear, disgust, neutral & Surprised
* The data is read from its respective folders and training and validation has to be performed respectively.

**References**

* <https://www.kaggle.com/datasets/noaa/hurricane-database>
* <https://www.kaggle.com/datasets/mysarahmadbhat/airline-passenger-satisfaction/data>

**Challenges**

* Limited Knowledge about the data sets and machine learning.

**Next Steps**

* Preparation for the data preprocessing
* More data searching
* Examine data through visualization and analysis techniques.
* Generate statistical summary
* Feature correlation to output class(es).
* Dealing with Categorical data
* Fixed problems like missing values, errors or outliers.
* Apply pre-processing or normalization procedures