**DATA PROGRAMMING**

**BDAT – 1004**

**FINAL PROJECT**

**SUBMITTED BY: SUBMITTED TO:**

GROUP – 4 PROFESSOR ROSHAN SAHU

KANIKA KHURANA (200514604)

PRINCE SINGLA (200552656)

SAMARJEET SINGH (200558169)

NAVNEET SINGH (200558447)

GURSHARAN SINGH SIBIA (200551485)

**PROJECT SUMMARY**

For this project we assumed the hypothetical client which is hotel group, **SUNNYSIDE INN**. So, there requirement from us is to select the real flowers everyday for their INN having different sizes based on the location they need to put in say for lobby they need bigger flowers than for the corridors, so that they can attract more guests to their hotel. So, for this project we have used the iris dataset from the "archive.ics.uci.edu" website which is the most similar dataset to the requirement which we need to fulfill for the hotel. This dataset contains data about the iris flower categorisation based on the length and width of sepals and petals. So, starting to visualize the dataset we clean it in order to ensure that the data is accurate, consistent and complete. After cleaning the dataset, we used matplotlib and seaborn to perform data visualisation to analyze the data and recognize patterns in it.

The dataset contains one column of categorical data (flower\_type) and 5 columns of the numeric type with non-null entries (flower\_attributes). The dataset contains no missing or null values. We also did not find any duplicate entries in dataset. It contains 149 rows of data about the sepal length, petal length, sepal width and petal width of 3 different flower types. We used pandas to acquire knowledge flower species.

Using the scatterplot, we found out that Species Setosa has smallest sepal lengths but larger sepal widths. Versicolor Species lies in the middle of the other two species in terms of sepal length and width. Species Virginica has larger sepal lengths but smaller sepal widths. Species Setosa has smaller petal lengths and widths. Versicolor Species lies in the middle of the other two species in terms of petal length and width. Species Virginica has the largest of petal lengths and widths. Using the pairplot we found out that Setosa has the smallest of petals widths and lengths. It also has the smallest sepal length but larger sepal widths. Using histogram, we found out that the highest frequency of the sepal length is between 30 and 35 which is between 5.5 and 6. The highest frequency of the sepal Width is around 70 which is between 3.0 and 3.5. The highest frequency of the petal length is around 50 which is between 1 and 2. The highest frequency of the petal width is between 40 and 50 which is between 0.0 and 0.5. Using Distplot, we found out that in the case of Sepal Length, there is a huge amount of overlapping. In the case of Sepal Width also, there is a huge amount of overlapping. In the case of Petal Length, there is a very little amount of overlapping. In the case of Petal Width also, there is a very little amount of overlapping. Using Boxplot, we found out that Species Setosa has the smallest features and less distributed with some outliers. Species Versicolor has the average features. Species Virginica has the highest features. Using Heatmap we found out that Petal width and petal length have high correlations. Petal length and sepal width have good correlations. Petal Width and Sepal length have good correlations.

This is the summary of the cleaning and visualization of the dataset.

The knowledge we gain from this project how to understand the client requirements and work accordingly.