### **Course 2 - IITK AIML Core: Applied Data Science with Python**

**Submitted by: Saurav Ganguly**

**Project - Sales Analysis**

The Sales Analysis Project is developed as a course end project of Course 2 (IITK AIML Core: Applied Data Science with Python).

**Brief snapshot about the project and the analysis used:**

1. The Data Analysis has been conducted on the data provided in the csv file in the project material, for the Australian based company AAL.
2. For the analysis of the following python libraries have been used:

*Numpy, pandas, matplotlib, seaborn, statistics, scipy etc.*

1. The csv file has been imported and converted into a DataFrame and the calculations and analysis are done on that.
2. The Analysis has been conducted using the following steps:

* Data Wrangling - Cleaning and sanitizing the data, making it free from any missing values. Standardization, Normalization and Winsorization methods have been applied on various columns to transform the DataFrame data into consumable information for time based analysis.
* Data Analysis - The data has been transformed and grouped into several aggregation to conduct analysis on several features. All the analyses are supported with proper seaborn plots for visualization of the analysis outputs along with deduction notes for each of the analyses.

The following analysis have been performed on the DataFrame:

* Descriptive analysis on the whole data to calculate mean, median, mode, standard deviation, variance, percentile, covariance, correlation, skewness analysis etc.
* Descriptive analysis on the Weekly, Monthly and Quarterly data.
* Total and average sales analysis for all the Groups, to show the sales trend for all the groups along with outlier detection
* Total and average sales analysis for all the States, to show the sales trend for all the States along with outlier detection
* Total and average sales analysis for Time-of-Day, to show the sales trend for all the time of the day along with outlier detection
* Weekly Sales Analysis has been done for the below groups:
  + Total Weekly Sales Analysis
  + Average Weekly Sales Analysis
  + Weekly Sales distribution with outliers
  + Weekly Group-wise total sales
  + Weekly Group-wise average sales
  + Weekly Group-wise sales distribution with outliers
  + Weekly State-wise total sales
  + Weekly State-wise average sales
  + Weekly State-wise sales distribution with outliers
  + Weekly Time-wise total sales
  + Weekly Time-wise average sales
  + Weekly Time-wise sales distribution with outliers
* Monthly Sales Analysis has been done for the below groups:
  + Total Monthly Sales Analysis
  + Average Monthly Sales Analysis
  + Monthly Sales distribution with outliers
  + Monthly Group-wise total sales
  + Monthly Group-wise average sales
  + Monthly Group-wise sales distribution with outliers
  + Monthly State-wise total sales
  + Monthly State-wise average sales
  + Monthly State-wise sales distribution with outliers
  + Monthly Time-wise total sales
  + Monthly Time-wise average sales
  + Monthly Time-wise sales distribution with outliers
* Quarterly Sales Analysis has been done for the below groups:
  + Total Quarterly Sales Analysis
  + Average Quarterly Sales Analysis
  + Quarterly Sales distribution with outliers
  + Quarterly Group-wise total sales
  + Quarterly Group-wise average sales
  + Quarterly Group-wise sales distribution with outliers
  + Quarterly State-wise total sales
  + Quarterly State-wise average sales
  + Quarterly State-wise sales distribution with outliers
  + Quarterly Time-wise total sales
  + Quarterly Time-wise average sales
  + Quarterly Time-wise sales distribution with outliers.
* State-wise sales analysis for different demographic groups (kids, women, men, and seniors)
* Group-wise sales analysis (Kids, Women, Men, and Seniors) across various states
* Time-of-the-day analysis to identify peak and off-peak sales periods
* Checking Skewness of Data
  + Skewness of the Sales Distribution
  + Log Normal Distribution of Sales
  + Normalized Sales Distribution
  + Sqrt\_Sales(Square Root Sales) Distribution
  + Winsorized Sales Distribution
  + Skewness of the Unit Distribution
  + Log Normal Distribution of Units
  + Normalized Unit Distribution
  + Sqrt\_Unit(Square Root Unit) Distribution