**SALES EFFECTIVENESS**

**CATEGORY: PRODUCT SALES**

**Project Ref: PM-PR-0019**

**INTRODUCTION: -**

FicZon Inc is an IT solution provider with products ranging from on-premises products to SAAS based solutions. FicZon major leads generation channel is digital and through their website.

FicZon business is majorly dependent on the sales force effectiveness. As the market is maturing and more new competitors are entering the market, FicZon is experiencing a dip in sales.

Effective sales are dependent on lead quality and as of now, this is based on manual categorization and highly dependent on sales staff. Though there is a quality process, which continuously updates the lead categorization, its value is in for post analysis, rather than conversation.

FicZon wants to explore Machine Learning to pre-categorize the lead quality and as result, expecting significant increase in sales effectiveness.

**PROJECT** **GOALS: -**

1. Data exploration insights – Sales effectiveness.

2. ML model to predict the Lead Category (High Potential, Low Potential)

**DATA** **SOURCE: -**

Data is collected from the Client Database

DB Name: project\_sales

Table Name: data

Host: 18.136.157.135 Port: 3306

Username: dm\_team2

**FEATURE** **DETAILS: -**

Data columns (total 9 columns):

Created: - Gives the information about the data when it was created, particularly date and time.

Product\_ID: - It gives the unique identification number of the product.

Source: - It gives information through which platform the product is purchased like website, call, CRM form, etc....

Mobile: - It contains customers' contact numbers.

EMAIL: - It contains customer email.

Sales\_Agent: - It shows which sales agent was responsible for the sale of product.

Delivery\_Mode: - It is the which made of delivery.

The above features are the input variables.

Status: - It gives the leads of product sales. And it is the target variable.

**REQUIEMENTS: =**

Required Libraries are as follows:

Pymysql : Used to connect to the database.

NumPy: NumPy offers comprehensive mathematical functions, random number generator and more.

Pandas: Pandas are an open-source python package that is most widely used for data science and data analysis.

Matplotlib: Matplotlib is a plotting library for python programming language and its numerical extension NumPy.

Seaborn: Seaborn is a python data visualization library based on matplotlib. It provides high level interface for drawing attractive and informative statistical graphics.

Warnings: Warnings are provided to warn the developer of situations that aren’t necessarily exceptions.

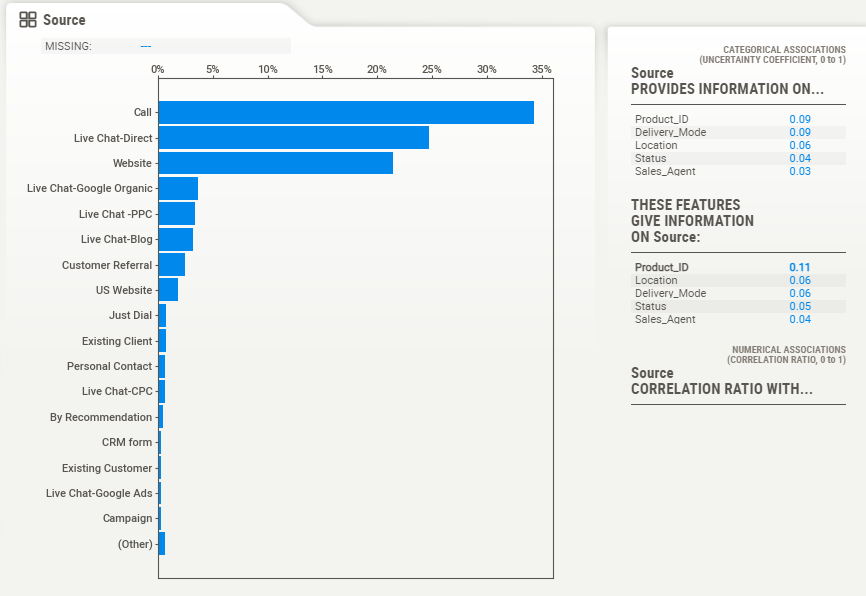
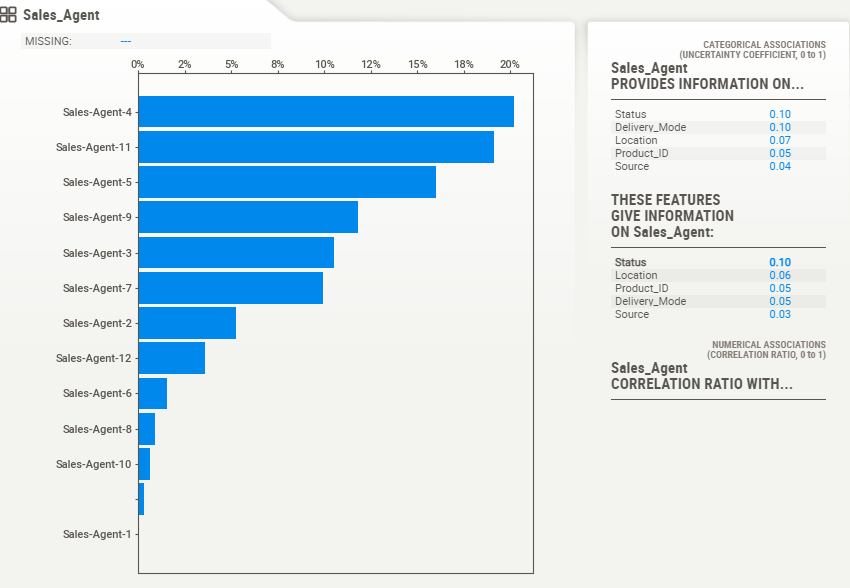
**BASIC INFORMATION: -**

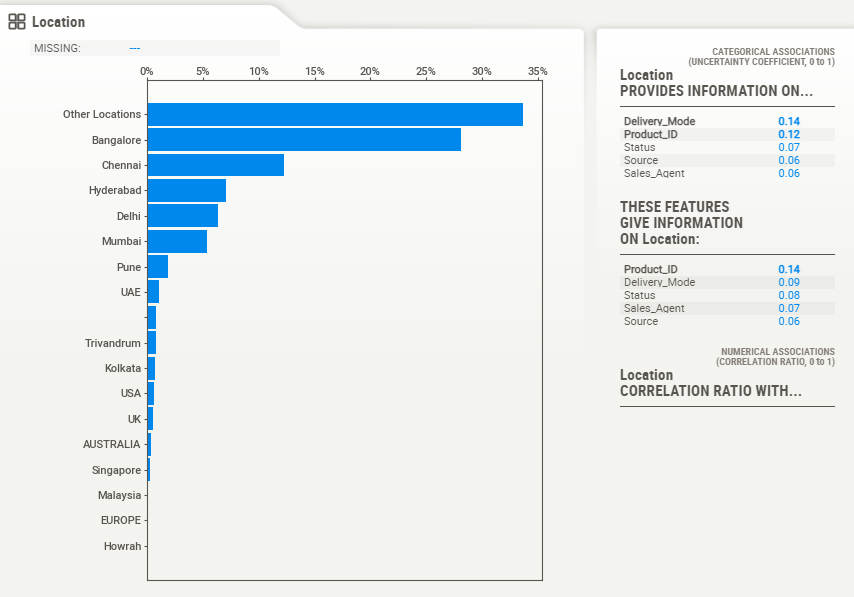
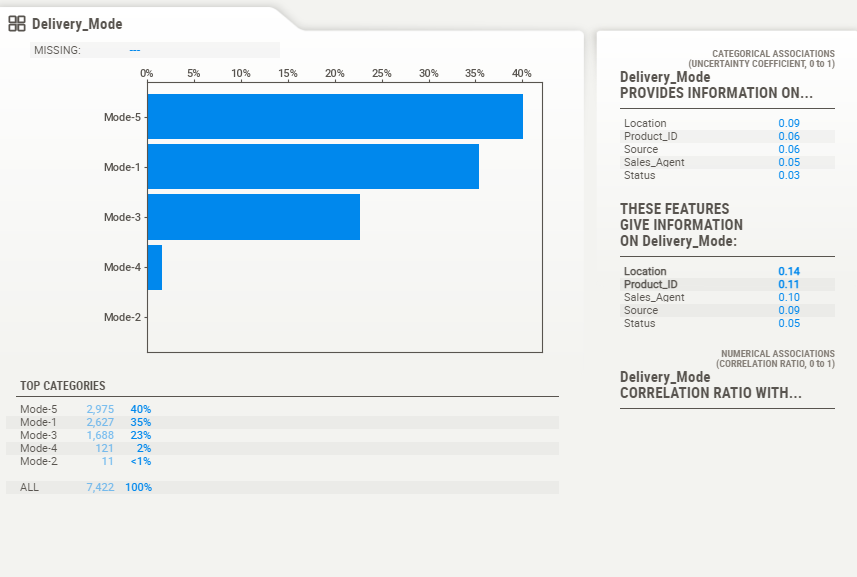
Basic information about the data:

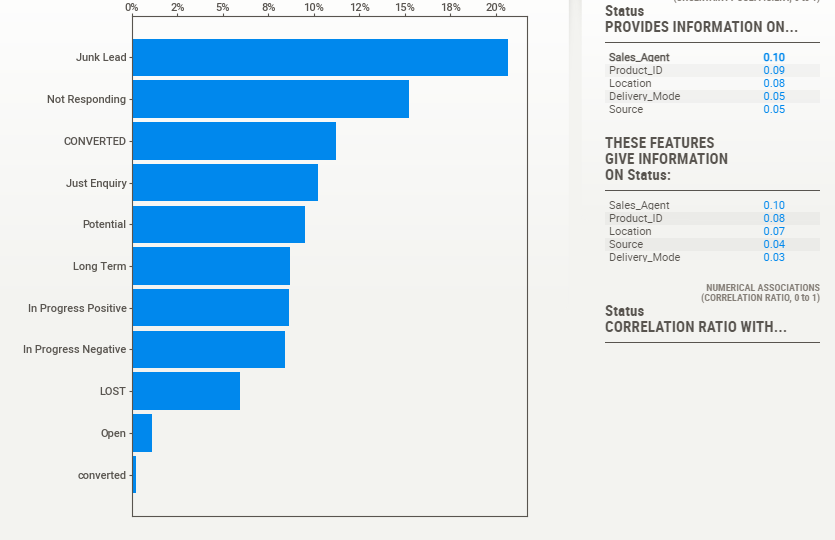
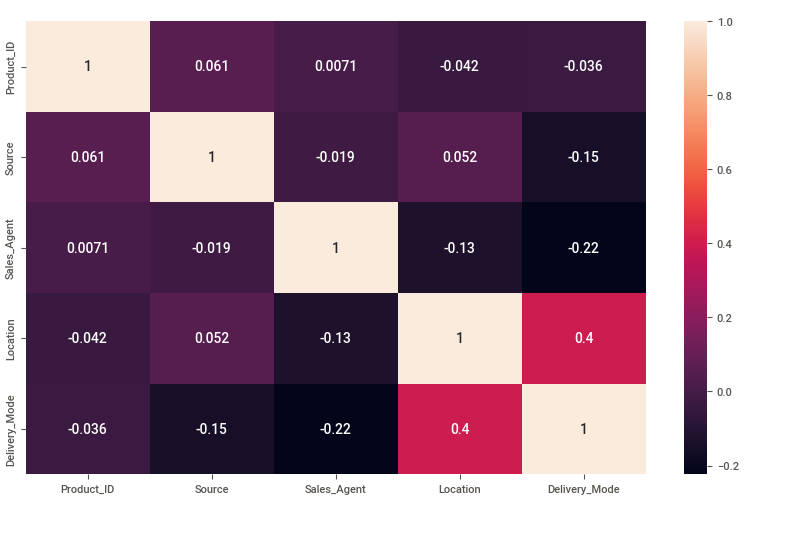
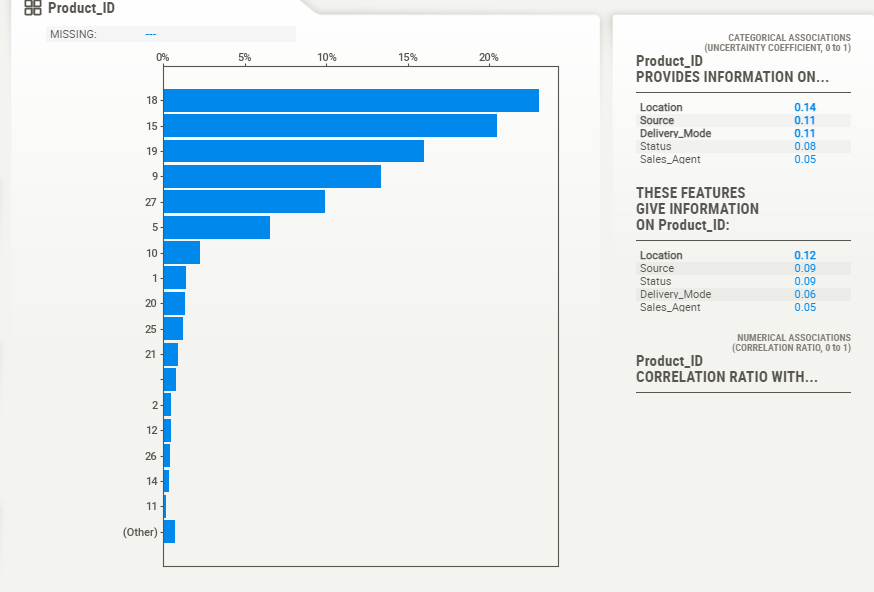
Data contains 7422 rows and 9 columns. In column Created contain 6752 Unique values, in column Product\_ID contain 30 Unique values, in column Source contain 26 unique values, in column Mobile 487 unique values, in column Email contain 883 unique values, in column Sales\_Agent contains 13 unique values, in column Location 18 unique values, in Delivery\_Mode contains 5 unique values, in Status contains 11 unique values. Top product sales were recorded on 26-09-2018 and it was a junk lead. The data has no missing values, no duplicate values.

**Analysis: -**

Sales of product is High on calls, live chat direct and website. Sales agent 4, sales agent 11 has made the highest sales. Bangalore has the highest sales of the product. Products have done highest delivery through mode 5. Sales have been made majorly through junk leads. Product ID 18 and 15 selling highly. The information above is shown graphically in the below figures.

**MODEL CREATION: -**

Creating models:

Firstly, we have converted categorical data into numerical data to get an accurate model. We have given Source, Sales Agent, location, delivery mode as training of the model Status as taken as testing of the model. XGB Classifier is a scalable, distributed gradient boosted decision tree. It provides parallel tree boosting and is the leading machine learning library for regression, classification and ranking problems. It got 100% accuracy. Random Forest is a powerful and versatile supervised machine learning algorithm that grows and combines multiple decision trees to create a forest. It got 100% accuracy. KNN is a non-parametric, supervised learning classifier, which uses proximity to make classifications or predictions about the grouping of an individual data point. It got 96% accuracy.

**CONCLUSION: -**

### Machine Learning can predict to pre-categorize the lead quality and as result. We are confident this model can help client to increase sales effectiveness. Accuracy of model 100%, we can use Random Forest or XGB Classifier. But for KNN Classifier, we cannot recommend it, because the score is below 100%.