**SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS**

*A project report submitted to ICT Academy of Kerala*

*in partial fulfillment of the requirements*

*for the certification of*

**CERTIFIED SPECIALIST**

**IN**

**DATA SCIENCE & ANALYTICS**

submitted by

**Aneete Rose K.S**

**Megha S.S**

**Kavitha K**

**Nitheesh R.S**



**ICT ACADEMY OF KERALA**

**THIRUVANANTHAPURAM, KERALA, INDIA**

**Feb 2022**

**List of Figures**

**Table of Contents**

**Abstract**

In recent years, the technologies have been unbelievable and more companies are moving towards the adoption of these technologies to share information in real-time which leads to the generation of a huge amount of data. This data, when used correctly, will be very helpful to the company to discover hidden patterns for better decision making in the future. Here we are working on a project named ‘SMART SUPPLY CHAIN FOR BIG DATA ANALYSIS’. Since the dataset used is related to supply chain, important parameters are identified and the machine learning models are trained with the dataset for detection of fraud transactions, late delivery of orders, sales revenue and quantity of products which customer orders. This project aims to compare different machine learning models and measure their performance to find out which machine learning model performs better.

**1. Problem Definition**

**1.1 Overview**

A Dataset of Supply Chains used by the company Data Co Global was used for the analysis. Dataset of Supply Chain, which allows the use of Machine Learning Algorithms and Python software.

**1.2 Problem Statement**

# Sales prediction and Fraud detection on Supply Chain Dataset

**2. Introduction**

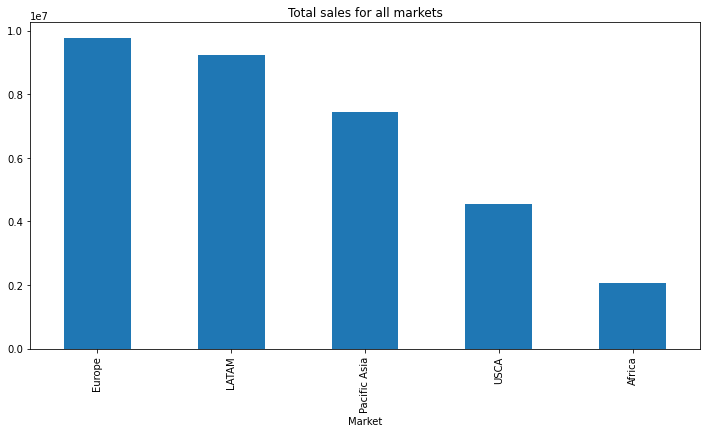
This is about the analysis made on the dataset obtained from, DataCo supply chain.

The variables in this dataset can be related to the supply chain around an e-commerce company. In specific, the variables contain Customer, Production, Sales, Stores, and Logistics information.

Analysis is carried out by using the data visualization techniques to uncover patterns and generate insights.

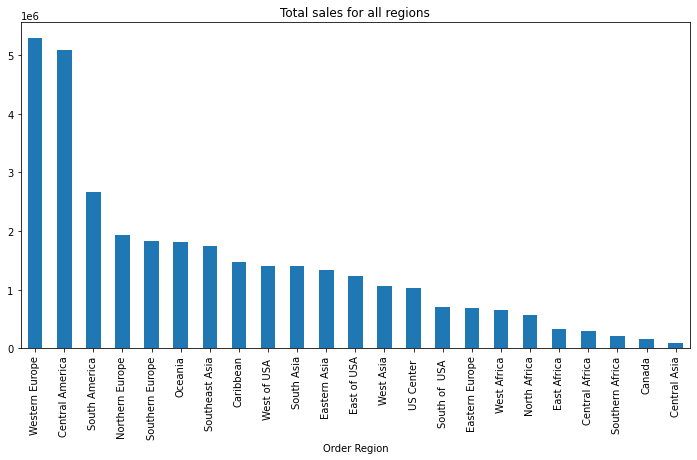
**7. Result**

Fig 7.1



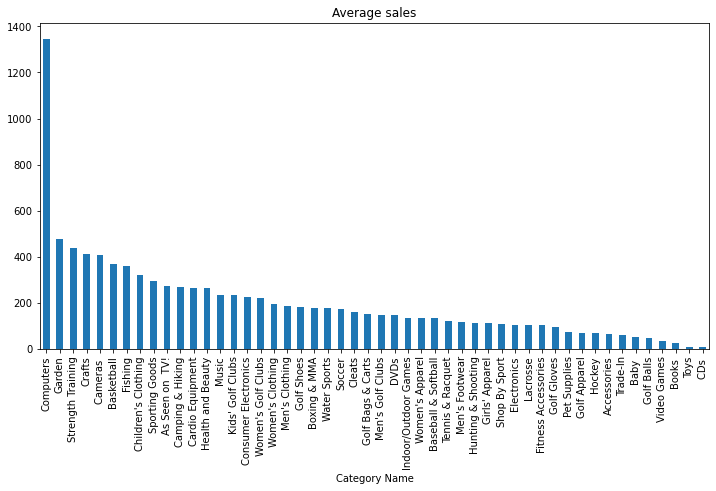
By the analysis of sales in each market and region using bar chart, we understood that European market and western European region have the most sales and Africa and central Asia has least sales.

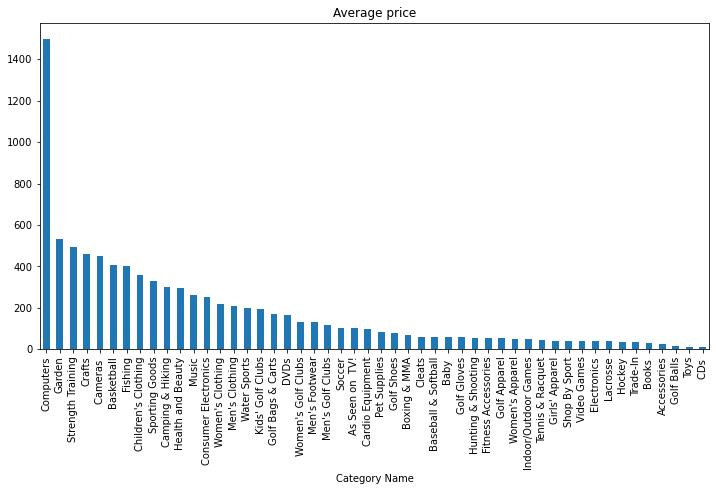
Fig 7.2



In region wise, Western Europe came first and Central Asia came last in sales.

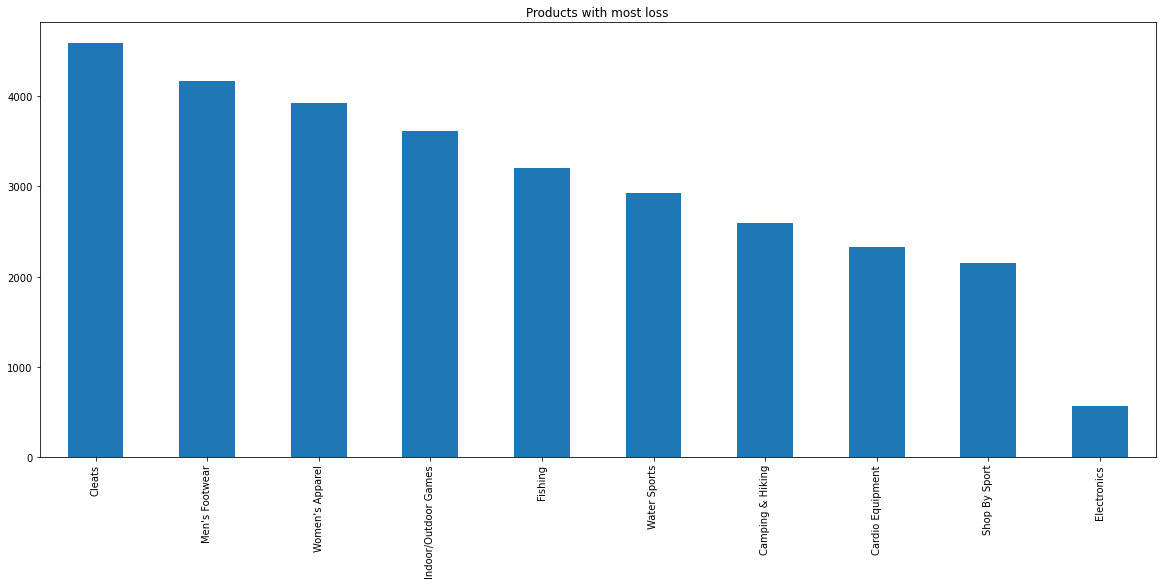
Fig 7.3

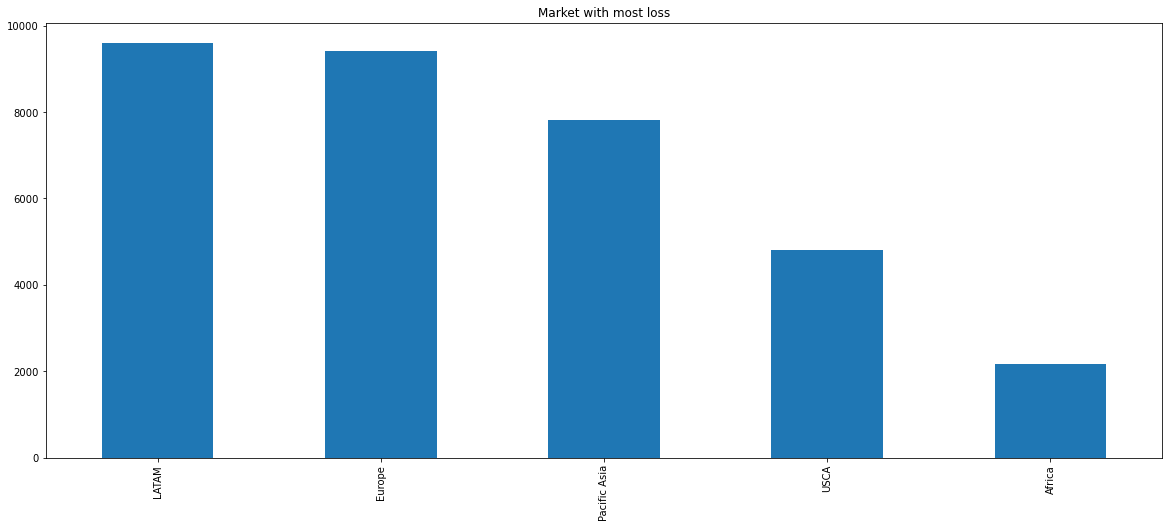
****

****

Some of the top products with highest price on average are the most sold products on average.

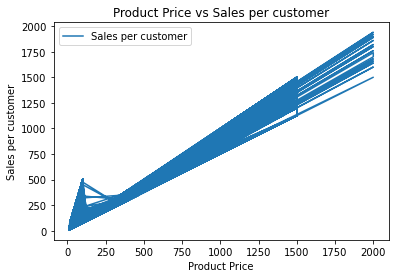
Fig 7.4

****

****

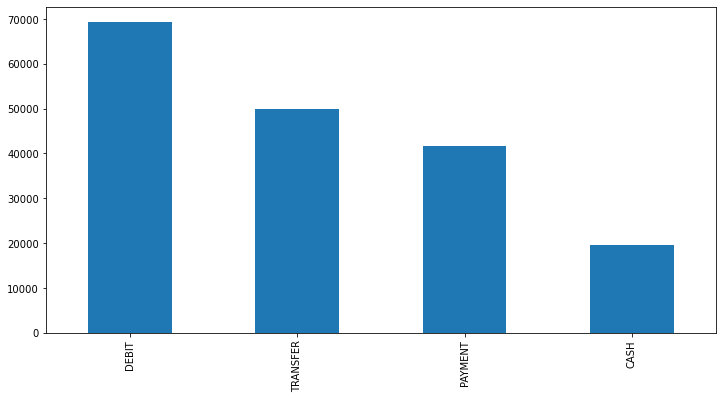
Product with most loss is cleats and Market with most loss is Latin America

Fig 7.5

****

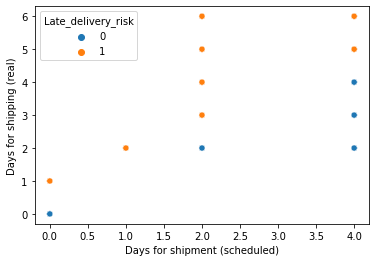
sales and product price have a linear relationship with each other.

Fig 7.6

****

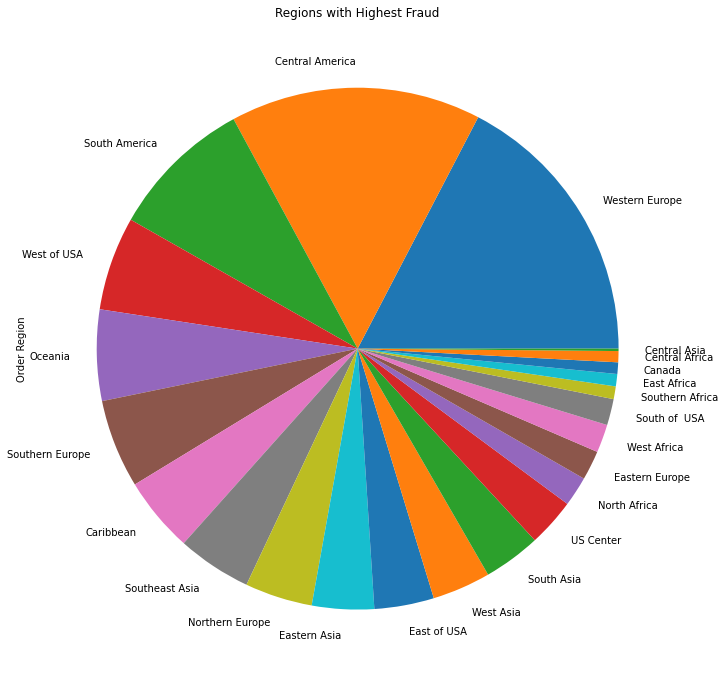
Debit type is the most preferred payment method and cash payment is the least preferred method.

Fig 7.7

****

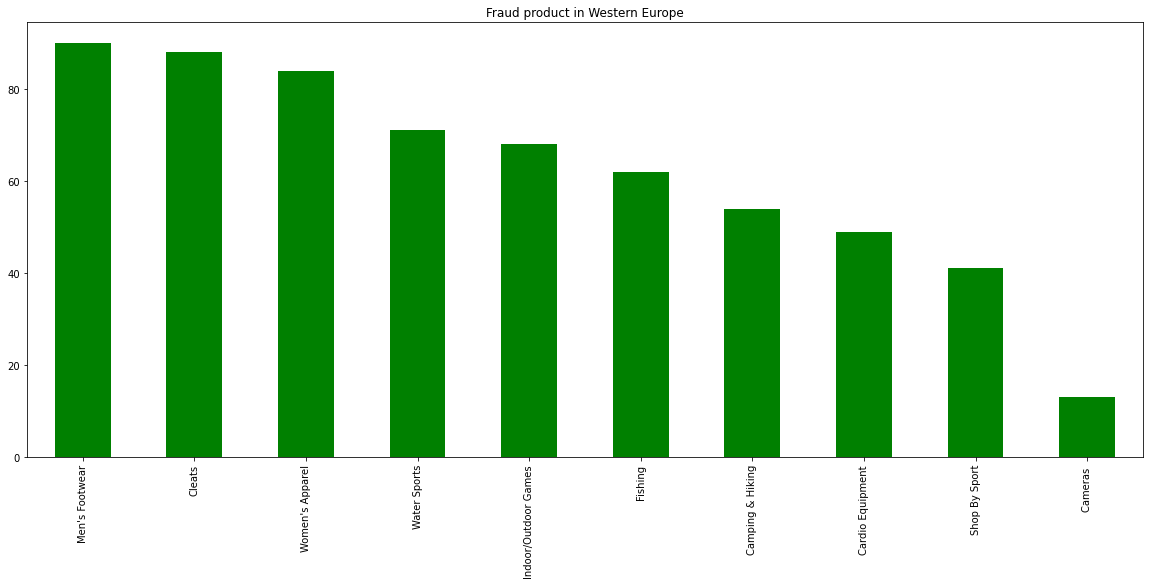
From this we can understand the risk of late delivery related to days for scheduled shipment and real.

Fig 7.8

****

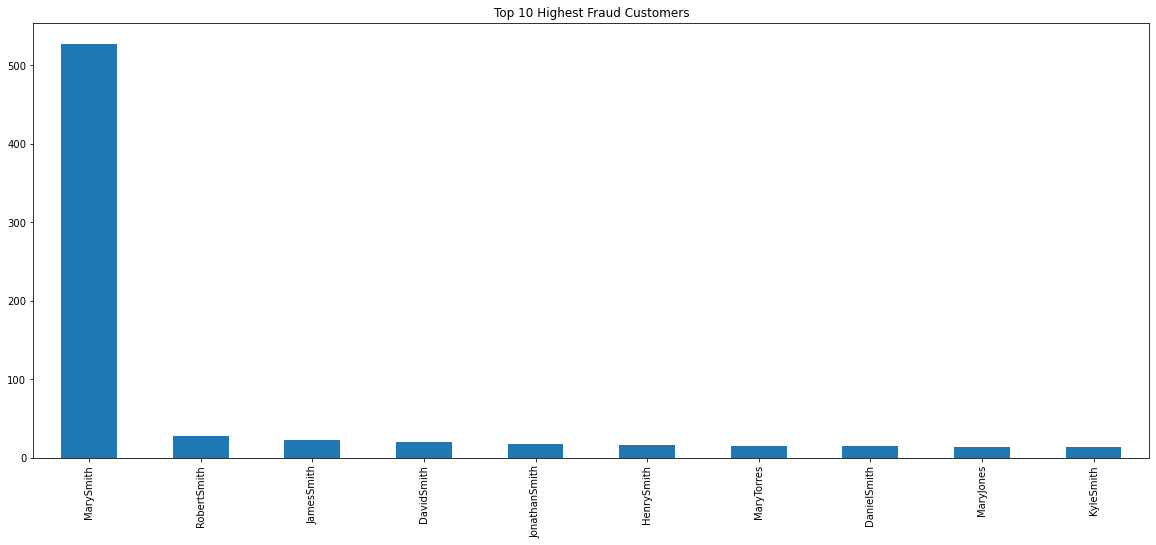
Highest number of suspected fraud orders are from Western Europe and Central America.

Fig 7.9



Men's Footwear and Cleats products are the most suspected fraud orders in Western Europe.

Fig 7.10



The customer named Mary Smith was responsible for most of the fraud transactions.