

**A**

**Project Report**

**On**

**“**Automobile Spare Parts Management System**”**

**Submitted to**

**Bharati Vidyapeeth**

**(Deemed to be University ) , Pune**

**In partial fulfillment of the requirement of**

**MASTER OF COMPUTER APPLICATION (MCA-III ,SEM-VI)**

**2020-21**

**Submitted By**

**Mr. Ganraj Rajendra Dol.**

**Under the Guidance of**

**Dr. A. V. Nikam**

**Through**

**THE DIRECTOR**

**Bharati Vidyapeeth**

**(Deemed to be University ) , Pune**

**Yashwantrao Mohite Institute of Management, Karad**

**2020-21**

****

**Bharati Vidyapeeth**

**(Deemed to be University ) , Pune**

**Yashwantrao Mohite Institute of Management, Karad**

**CERTIFICATE**

This is certify that project entitled **“Automobile Spare Parts Management System”**  Submitted by **Mr. Dol Ganraj Rajendra.** in partial fulfilment of the requirement for award of **”MASTER OF COMPUTER APPLICATION(SEM-VI)”** has submitted his work satisfactory .

To the best of my knowledge and belief , the matter presented in practical assignment have not been submitted earlier.

**Date :   / /2021**

**Place : Karad .**

**Director**

**(Dr. Rajesh Kanthe)**

****

**Bharati Vidyapeeth**

**(Deemed To Be University ) , Pune**

**Yashwantrao Mohite Institute of Management, Karad**

**CERTIFICATE**

This is certify that project entitled **“Automobile Spare Parts Management System”**  Submitted by **Mr. Dol Ganraj Rajendra** in partial fulfilment of the requirement for award of **”MASTER OF COMPUTER APPLICATION(SEM-VI)”** has submitted his work satisfactory .

To the best of my knowledge and belief , the matter presented in practical assignment have not been submitted earlier.

**Date :   / /2021**

**Place : Karad .**

**Project Guide     Head of Department**

**(Dr. A.V. Nikam) (Dr. R. J. Jadhav)**

****

**Bharati Vidyapeeth**

**(Deemed to be University ) , Pune**

**Yashwantrao Mohite Institute of Management, Karad**

**CERTIFICATE**

This is certify that project entitled **“Automobile Spare Parts Management System”** Submitted by **Mr. Patil Atul Ganapati , Mr. Dol Ganraj Rajendra** in partial fulfilment of the requirement for award of **”MASTER OF COMPUTER APPLICATION(SEM-III)”** has submitted his work satisfactory .

To the best of my knowledge and belief , the matter presented in practical assignment have not been submitted earlier.

**Date :   / /2021**

**Place : Karad .**

**External Examiner**

**DECLARATION**

To,

The Director,

Bharati Vidyapeeth

(Deemed to be University), Pune

Yashwantrao Mohite Institute Of Management,

Malkapur, karad.

Respected Sir,

We here by declare that the project report entitled **“*Automobile Spare Parts Management System*”** Written and submitted under the guidance of **Dr. A. V. Nikam** is our genuine record of own work. The matter collected in this report is not reproduction from any source.

We understand that, any such copying is liable to be punish by the rules of the university.

Date:- / /2021

Place:- Karad

Mr.Ganraj Rajendra Dol.

**Index**

|  |  |  |
| --- | --- | --- |
| Sr. No. | Particulars | Pages |
| 1 | Introduction to System | 7 |
| 2 | Information About Modules | 8-12 |
| 3 | Feasibilty Study | 12 |
| 4 | Introduction to Python | 14-16 |
| 5 | Introduction to Data Science | 16-21 |
| 6 | Information About DataSets | 22-24 |
| 7 | Diagrams | 25 |
| 8 | Source Code | 26 |
| 9 | OutPut | 26-36 |
| 11. | Conclusion | 37 |
| 12. | Bibliography | 38 |

**Overview**

* 1. **Introduction:**

Automobile Spare Parts Management System Project is implemented for managing the dealer, customers, claims, Parts, Transactions and also This system also provides an the company to check or analyze the part is in warranty or not and also provides company to check the Monthly report of Dealers and parts that has been failed &Predict the parts that may fail according to the location, mileage and age. Company to understand Monthly report of top ten failed parts with the help of this system they can check Quality of the parts, Pattern of failure mostly reported failed part in the quarter/year. Predict the parts that may fail according to the location, mileage and age: company can analyzing the historical data, we need to predict the parts that can be expected to fail according to the age, km driven and location of the vehicle. This data helps the company to understand the manufacturing requirements of the part i.e. which part is to be manufactured in what quantity so as to avoid unavailability of the parts when claim is raised for that part. To determine whether the defective part is to be replaced or repaired. The model was developed to decide whether the failed part has to be repaired or replaced with the new one. For this, we had the last claimed data of the company. As this is the binary classification problem, Decision tree and Random Forest algorithms are used to build the model. Customer Retention Model. This model helps the company to understand the service provided by the dealers is satisfactory or not. The model analyses the performance of the dealers by considering the customers feedback, churn rate of that location, preference to the other locations by the customers, change of base location etc. The model helps the company to take some regulatory decisions based the performance of dealer. It also helps the company to decide, whether to introduce new schemes, offers etc. to the affected location or to change the dealer. Processing of variety of data. The data were coming from various locations like US, Canada, China, and India in huge amount with varying data-types it was required to club the data and then process it further for the analyzing. Processing large volume of data also caused memory issues which were then handled by memory management

Automobile Spare Parts Management System is an application using [Google Co-laboratory](https://colab.research.google.com/notebooks/) in Python with data science. The feature of this application includes On the sale of the vehicle company provides one year base warranty on its parts and also provides optional one year extended warranty on additional cost. If the customer has not opted for the extended warranty option then after the one year in case of any failure of any part, the warranty is said to be expired and all the expenses has to be incurred by the customer itself and company will not take any responsibility for the same. On the other hand, if the vehicle is secured by extended warranty assurance, then after the one year from the sale i.e. after the base warranty period, if any failure happens then the affected part/parts is/are either repaired or replaced under the warranty terms of the company and all the expenses will be done by the company. This whole system is built using Python with data science programming language with the help of [Google Co laboratory](https://colab.research.google.com/notebooks/).

**Existing System**:

At present traders maintain their day to day company has to face many issue of any failure happens then the affected part/parts is/are either repaired or replaced under the warranty terms of the company and all the expenses will be done by the company. During this process it takes lots of time and dealer must wait till the process, only single system is used during this process.

The existing system is Manual. As the existing system is manual so accessibility of the record is time consuming . It is difficult to search the record throughout the system. As the system is very complex it is difficult to maintain database . Time required for searching and processing of record is more

**Proposed System:**

The Proposed System for Automobile Spare Parts Management System is to be designed in such a manner that it eliminates all communication and coordination barriers and also reduces shortcomings and limitations of existing systems.

The objectives of the program should be to ensure confidentiality, integrity and availability of information pertaining to the organization campus operations. The critical pieces of information that need protection from modification and disclosure data application. In addition to the information security objectives, the assurance program should aim to accomplish the organization goals.

**Modules :-**

**Exploratory Data Analysis**

Exploratory Data Analysis refers to the critical process of performing initial investigations on data so as to discover patterns, to spot anomalies, to test hypothesis and to check assumptions with the help of summary statistics and graphical representations. (For this example pandas, numpy, matplotlib and seaborn) and loaded the data set. EDA is an approach to analyze the data and find patterns, visual insights, etc. that the data set is having, before proceeding to model. One spends a lot of time doing EDA to get a better understanding of data

* **Data Quality Check:** Can be done using pandas library functions like describe(), info(), dtypes(), etc. It is used to find several features, its datatypes, duplicate values, missing value, etc.
* **Statistical Test:**Some statistical tests like Pearson correlation, Spearman correlation, Kendall test, etc are done to get a correlation between the features. It can be implemented in python using the *stats*library.
* **Quantitative Test:** Some quantitative test is used to find the spread of numerical features, count of categorical features. It can be implemented in python using the functions of the pandas library.
* **Visualization:**Feature visualization is very essential to get an understanding of the data. Graphical techniques like bar plots, pie charts are used to get an understanding of categorical features, whereas scatter plots, histograms are used for numerical features.

**Cross Validation**

## Validation

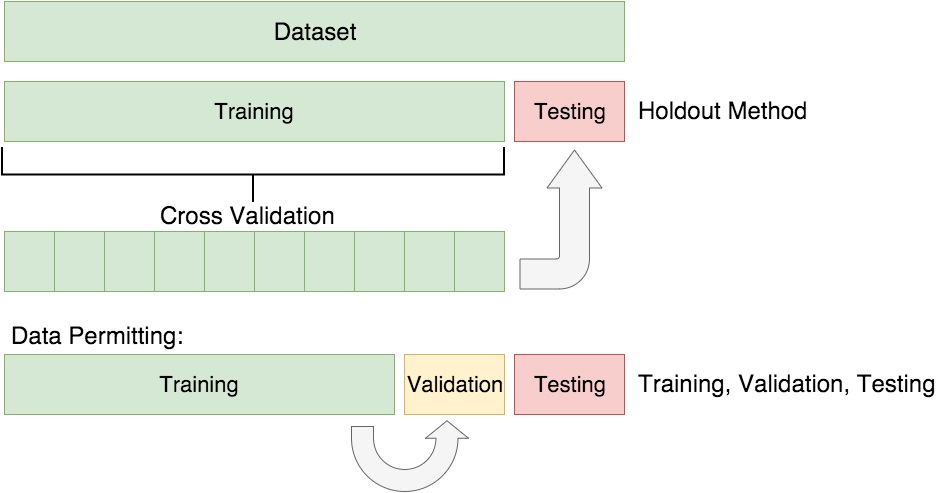
This process of deciding whether the numerical results quantifying hypothesized relationships between variables, are acceptable as descriptions of the data, is known as **validation**. Generally, an error estimation for the model is made after training, better known as evaluation of residuals. In this process, a numerical estimate of the difference in predicted and original responses is done, also called the training error. However, this only gives us an idea about how well our model does on data used to train it. Now its possible that the model is underfitting or overfitting the data. So, the **problem with this evaluation technique is that it does not give an indication of how well the learner will generalize to an independent/ unseen data set**. **Getting this idea about our model is known as Cross Validation.**

## Holdout Method

Now a basic remedy for this involves removing a part of the training data and using it to get predictions from the model trained on rest of the data. The error estimation then tells how our model is doing on unseen data or the validation set. **This is a simple kind of cross validation technique, also known as the holdout method**. Although this method doesn’t take any overhead to compute and is better than traditional validation, **it still suffers from issues of high variance**. **This is because it is not certain which data points will end up in the validation set and the result might be entirely different for different sets.**

## K-Fold Cross Validation

As there is never enough data to train your model, removing a part of it for validation poses a problem of underfitting. **By reducing the training data**, **we risk losing important patterns/ trends in data set, which in turn increases error induced by bias.** So, what we require is a method that provides ample data for training the model and also leaves ample data for validation. K Fold cross validation does exactly that.



## Variations on Cross-Validation

There are a number of variations on the k-fold cross validation procedure.

Three commonly used variations are as follows:

**Train/Test Split**: Taken to one extreme, k may be set to 2 (not 1) such that a single [train/test split](https://machinelearningmastery.com/train-test-split-for-evaluating-machine-learning-algorithms/) is created to evaluate the model.

**LOOCV**: Taken to another extreme, k may be set to the total number of observations in the dataset such that each observation is given a chance to be the held out of the dataset. This is called leave-one-out cross-validation, or [LOOCV](https://machinelearningmastery.com/loocv-for-evaluating-machine-learning-algorithms/) for short.

**Stratified**: The splitting of data into folds may be governed by criteria such as ensuring that each fold has the same proportion of observations with a given categorical value, such as the class outcome value. This is called [stratified cross-validation](https://machinelearningmastery.com/cross-validation-for-imbalanced-classification/).

**Repeated**: This is where the k-fold cross-validation procedure is [repeated n times](https://machinelearningmastery.com/repeated-k-fold-cross-validation-with-python/), where importantly, the data sample is shuffled prior to each repetition, which results in a different split of the sample.

**Nested**: This is where k-fold cross-validation is performed within each fold of cross-validation, often to perform hyperparameter tuning during model evaluation. This is called [nested cross-validation](https://machinelearningmastery.com/nested-cross-validation-for-machine-learning-with-python/) or double cross-validation.

## module dct crossvalidation.png

## Decision Tree Algorithm

A decision tree is a flowchart-like tree structure where an internal node represents feature(or attribute), the branch represents a decision rule, and each leaf node represents the outcome. The topmost node in a decision tree is known as the root node. It learns to partition on the basis of the attribute value. It partitions the tree in recursively manner call recursive partitioning. This flowchart-like structure helps you in decision making. It's visualization like a flowchart diagram which easily mimics the human level thinking. That is why decision trees are easy to understand and interpret.



Decision Tree is a white box type of ML algorithm. It shares internal decision-making logic, which is not available in the black box type of algorithms such as Neural Network. Its training time is faster compared to the neural network algorithm. The time complexity of decision trees is a function of the number of records and number of attributes in the given data. The decision tree is a distribution-free or non-parametric method, which does not depend upon probability distribution assumptions. Decision trees can handle high dimensional data with good accuracy.

## Decision Tree Classifier Building in Scikit-learn

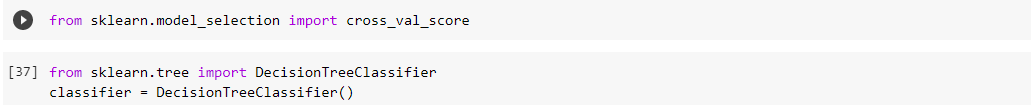
### Importing Required Libraries

Let's first load the required libraries.

decion tree libray.png

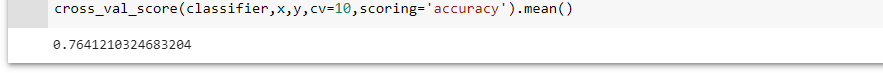
### Building Decision Tree Model

Let's create a Decision Tree Model using Scikit-learn.



### Evaluating Model

Let's estimate, how accurately the classifier or model can predict the type of cultivars.Accuracy can be computed by comparing actual test set values and predicted values.



**Feasibility Study**

The initial investigation points to the question whether the project is feasible. A feasibility is conducted to identify the best system that meets the all the requirements. This includes an identification description, a valuation of the proposed systems and selection of the best system for the job. The requirements of the system are specified with a set of constraints such as system objectives and the description of the out puts. It is then duty of the analyst to evaluate the feasibility of the proposed system to generate the above results. Three key factors are to be considered during the feasibility study.

* **Technical Feasibility -:**

The main consideration is to be given to the study of available resources of the organization where the software is to be implemented. Here the system analyst evaluates the technical merits of the system giving emphasis on the performance, Reliability, maintainability and productivity

By taking the consideration before developing the proposed system, the resources availability of the organization was studied. The organization was immense computer facilities equipped with sophisticated machines and the software hence this technically feasible.

* **Operational Feasibility -:**

An estimate should be made to determine how much effort and care will go into the developing of the system including the training to be given to the user. Usually, people are reluctant to changes that come in their progression. The computer initialization will certainly affected the turn over, transfer and employee job status. Hence an additional effort is to be made to train and educate the users on the new way of the system.

* **Economical Feasibility -:**

Economic feasibility is the most important and frequently used method for evaluating the effectiveness of the proposed system. It is very essential because the main goal of the proposed system is to have economically better result along with increased efficiency. Cost benefit analysis is usually performed for this purpose. It is the comparative study of the cost verses the benefit and savings that are expected from the proposed system. Since the organization is well equipped with the required hard ware, the project was found to be economically.

**System Analysis**:

- Hardware Requirements:

Processor : i3 processor

Memory : 4 GB RAM

Display : 14’’ LCD

Hard disk Drive : 120 GB

- Software Requirements : [Google Co-laboratory](https://colab.research.google.com/notebooks/), Jupyter notebook

Operating System : Windows 7, 8.1,10.

1.5.1 **Introduction to Python with data science:**

**Python** is a dynamic, object-oriented, high-level programming language that can be used for many kinds of software development. Python programming language was created by Guido van Rossum. Python got its name from the BBC comedy series “Monty Python’s Flying Circus”. Guido van Rossum needed a name that was short, unique, and slightly mysterious. Since he was a fan of the show he thought this name was great. Python is an interpreter, interactive, object-oriented programming language. It incorporates modules, exceptions, dynamic typing, very high level dynamic data types, and classes. It supports multiple programming paradigms beyond object-oriented programming, such as procedural and functional programming. Python combines remarkable power with very clear syntax. It has interfaces to many system calls and libraries, as well as to various window systems, and is extensible in C or C++. It is also usable as an extension language for applications that need a programmable interface. Python is portable: it runs on many UNIX variants including Linux and macOS, and on Windows.

**Features of Python Programming language:**

At present, there are many programming languages, and everyone has its own unique and different feature. It is the feature which matters the most and helps planning of choosing the best programming language for a project. So, you should know about the Python Programming language before you decide which is better for you. For knowing a language, you should know the feature, below are the features that will clear your thoughts about why Python over R or any other roots.

1. Easy to code and Read
2. High-Level Programming Language
3. Portable
4. Expressive
5. Object Oriented
6. Free and Open Source
7. Interpreted
8. Extensible
9. Embeddable
10. Large and Standard Library
11. GUI programming
12. Dynamically Typed

1. **Easy to code**:   
Python is a high-level programming language. Python is very easy to learn the language as compared to other languages like C, C#, JavaScript, Java, etc. It is very easy to code in python language and anybody can learn python basics in a few hours or days. It is also a developer-friendly language.

2. **Free and Open Source**:   
Since it is open-source; this means that source code is also available to the public. So you can download it as, use it as well as share it.

3. **Object-Oriented Language**:   
One of the key features of python is Object-Oriented programming. Python supports object-oriented language and concepts of classes, objects encapsulation, etc.

4. **GUI Programming Support**:   
Graphical User interfaces can be made using a module such as PyQt5, PyQt4, wxPython, or Tk in python.  
PyQt5 is the most popular option for creating graphical apps with Python.

5. **High-Level Language**:   
Python is a high-level language. When we write programs in python, we do not need to remember the system architecture, nor do we need to manage the memory.

6**. Extensible feature**:  
Python is a Extensible language. We can write us some Python code into C or C++ language and also we can compile that code in C/C++ language.

7. **Python is Portable language**:   
Python language is also a portable language. For example, if we have python code for windows and if we want to run this code on other platforms such as Linux, UNIX, and Mac then we do not need to change it, we can run this code on any platform.

8. **Python is Integrated language:**  
Python is also an Integrated language because we can easily integrated python with other languages like c, c++, etc.

9. **Interpreted Language**:  
Python is an Interpreted Language because Python code is executed line by line at a time. like other languages C, C++, Java, etc. there is no need to compile python code this makes it easier to debug our code. The source code of python is converted into an immediate form called byte code.

10. **Large Standard Library**  
Python has a large standard library which provides a rich set of module and functions so you do not have to write your own code for every single thing. There are many libraries present in python for such as regular expressions, unit-testing, web browsers, etc.

11. **Dynamically Typed Language**:  
Python is a dynamically-typed language. That means the type (for example- int, double, long, etc.) for a variable is decided at run time not in advance because of this feature we don’t need to specify the type of variable.

Apart from the above-mentioned features, Python has a big list of good features, few are listed below −

* It supports functional and structured programming methods as well as OOP.
* It can be used as a scripting language or can be compiled to byte-code for building large applications.
* It provides very high-level dynamic data types and supports dynamic type checking.
* It supports automatic garbage collection.
* It can be easily integrated with C, C++, COM, ActiveX, CORBA, and Java.

# Applications for Python

Python is used in many application domains.

* The Python Package Index lists thousands of third party modules for Python.

# Web and Internet Development

Python offers many choices for web development:

* Frameworks such as Django and Pyramid.
* Micro-frameworks such as Flask and Bottle.
* Advanced content management systems such as [Plone](http://www.plone.org/) and Django CMS.

Python's standard library supports many Internet protocols:

* HTML and XML
* JSON
* E-mail processing.
* Support for FTP, IMAP, and other Internet protocols.
* Easy-to-use socket interface.

And the Package Index has yet more libraries:

* Requests, a powerful HTTP client library.
* Beautiful Soup, an HTML parser that can handle all sorts of oddball HTML.
* Feedparser for parsing RSS/Atom feeds.
* Paramiko, implementing the SSH2 protocol.
* Twisted Python, a framework for asynchronous network programming.
* Microsoft Foundation Classes through the win32 extensions

**Data science**

Data science is the study of data. Like biological sciences are a study of biology, physical sciences, it’s the study of physical reactions. Data is real, data has real properties, and we need to study them if we’re going to work on them. Data Science involves data and some signs.

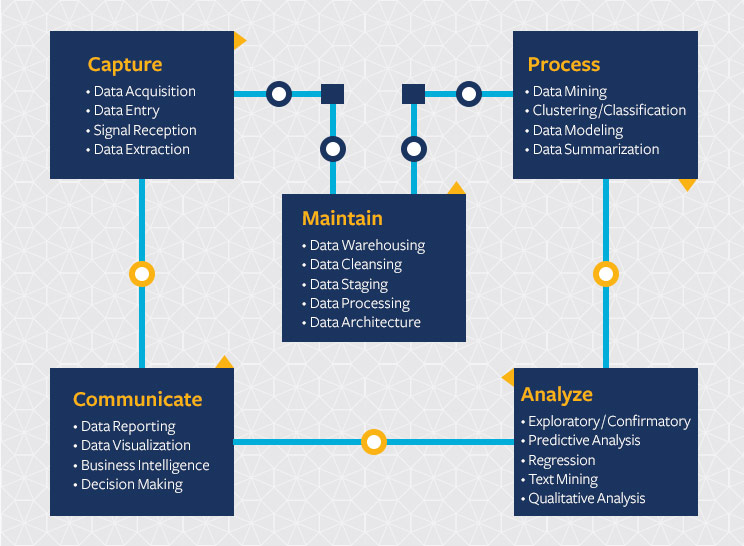
It is a process, not an event. It is the process of using data to understand too many different things, to understand the world. when you have a model or proposed explanation of a problem, and you try to validate that proposed explanation or model with your data.  
It is the skill of unfolding the insights and trends that are hiding (or abstract) behind data. It’s when you translate data into a story. So use storytelling to generate insight. And with these insights, you can make strategic choices for a company or an institution.

We can also define data science as a field which is about processes and systems to extract data of various forms and from various resources whether the data is unstructured or structured.

The definition and the name came up in the 1980s and 1990s when some professors, IT Professionals, scientist were looking into the statistics curriculum, and they thought it would be better to call it data science and then later on data analytics derived.

**Data science features and applications**

Data science continues to evolve as one of the most promising and in-demand career paths for skilled professionals. Today, successful data professionals understand that they must advance past the traditional skills of analyzing large amounts of data, data mining, and programming skills. In order to uncover useful intelligence for their organizations, data scientists must master the full spectrum of the data science life cycle and possess a level of flexibility and understanding to maximize returns at each phase of the process.

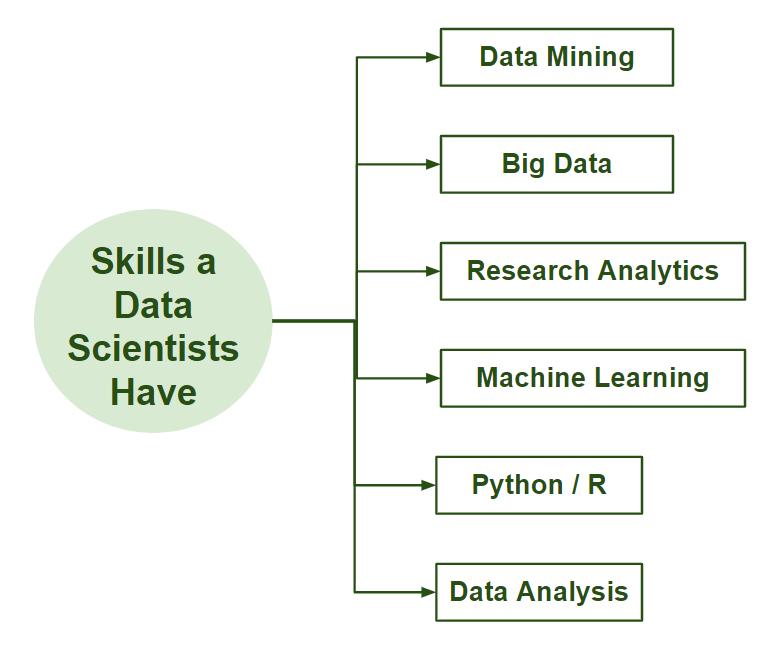


## ****Data Scientist****

There are several definitions available on Data Scientists. In simple words, a Data Scientist is one who practices the art of Data Science. The term “Data Scientist” has been coined after considering the fact that a Data Scientist draws a lot of information from the scientific fields and applications whether it is statistics or mathematics.

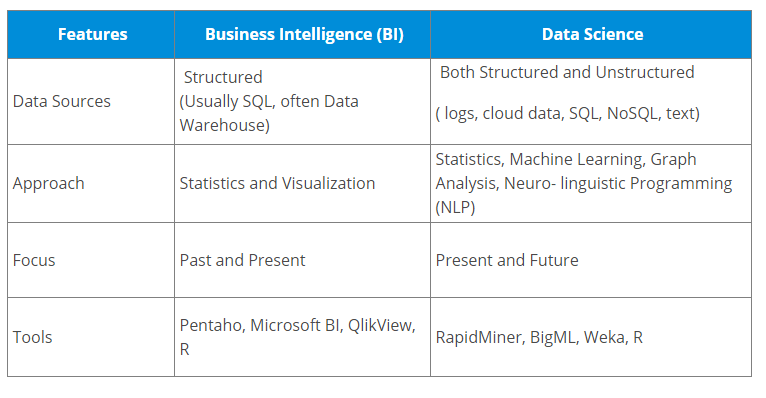
### ****What does a Data Scientist do?****

Data scientists are those who crack complex data problems with their strong expertise in certain scientific disciplines. They work with several elements related to mathematics, statistics, computer science, etc (though they may not be an expert in all these fields). They make a lot of use of the latest technologies in finding solutions and reaching conclusions that are crucial for an organization’s growth and development. Data Scientists present the data in a much more useful form as compared to the raw data available to them from structured as well as unstructured forms.



## Business Intelligence (BI) vs. Data Science

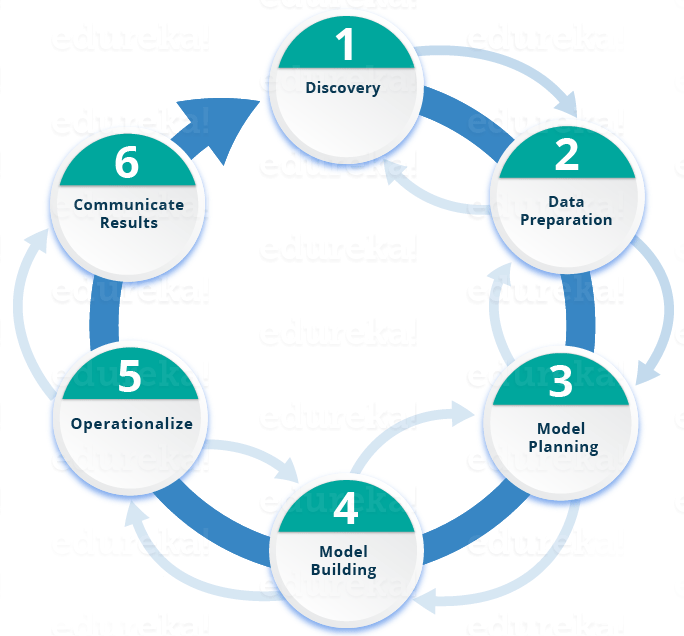
* Business Intelligence (BI) basically analyzes the previous data to find hindsight and insight to describe business trends. Here BI enables you to take data from external and internal sources, prepare it, run queries on it and create dashboards to answer questions like quarterly revenue analysis or business problems. BI can evaluate the impact of certain events in the near future.
* Data Science is a more forward-looking approach, an exploratory way with the focus on analyzing the past or current data and predicting the future outcomes with the aim of making informed decisions. It answers the open-ended questions as to “what” and “how” events occur.



A common mistake made in Data Science projects is rushing into data collection and analysis, without understanding the requirements or even framing the business problem properly. Therefore, it is very important for you to follow all the phases throughout the lifecycle of Data Science to ensure the smooth functioning of the project.

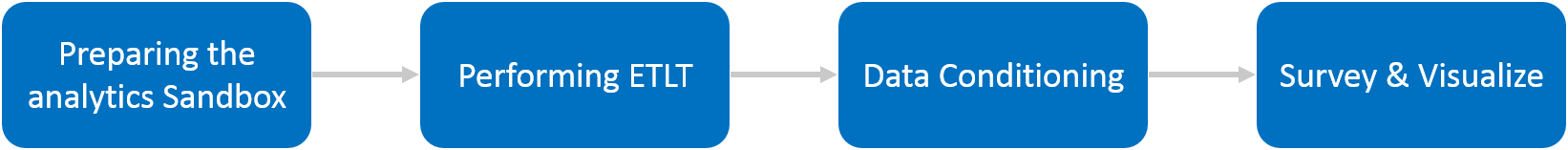
## Lifecycle of Data Science

Here is a brief overview of the main phases of the Data Science Lifecycle:



Phase 1—Discovery: Before you begin the project, it is important to understand the various specifications, requirements, priorities and required budget. You must possess the ability to ask the right questions. Here, you assess if you have the required resources present in terms of people, technology, time and data to support the project. In this phase, you also need to frame the business problem and formulate initial hypotheses (IH) to test.

 **Phase 2—Data preparation:**In this phase, you require analytical sandbox in which you can perform analytics for the entire duration of the project. You need to explore, preprocess and condition data prior to modeling. Further, you will perform ETLT (extract, transform, load and transform) to get data into the sandbox. Let’s have a look at the Statistical Analysis flow below.



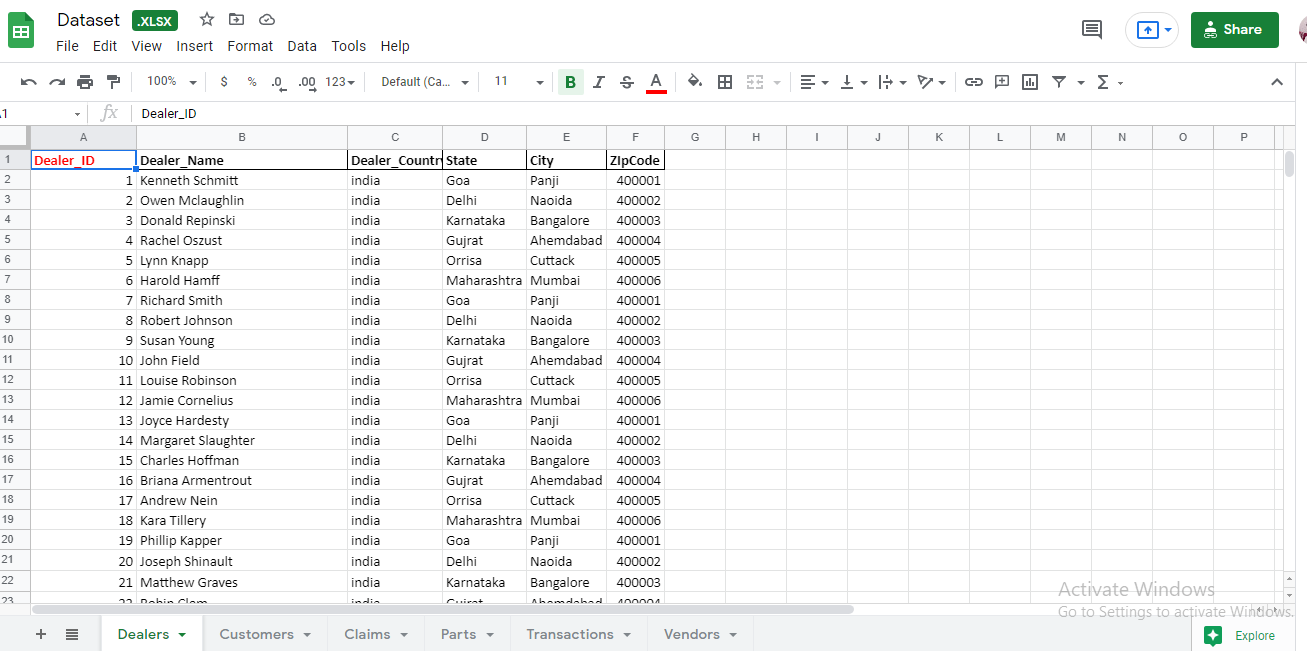
Phase 3—Model planning:  Here, you will determine the methods and techniques to draw the relationships between variables. These relationships will set the base for the algorithms which you will implement in the next phase. You will apply Exploratory Data Analytics (EDA) using various statistical formulas and visualization tools.

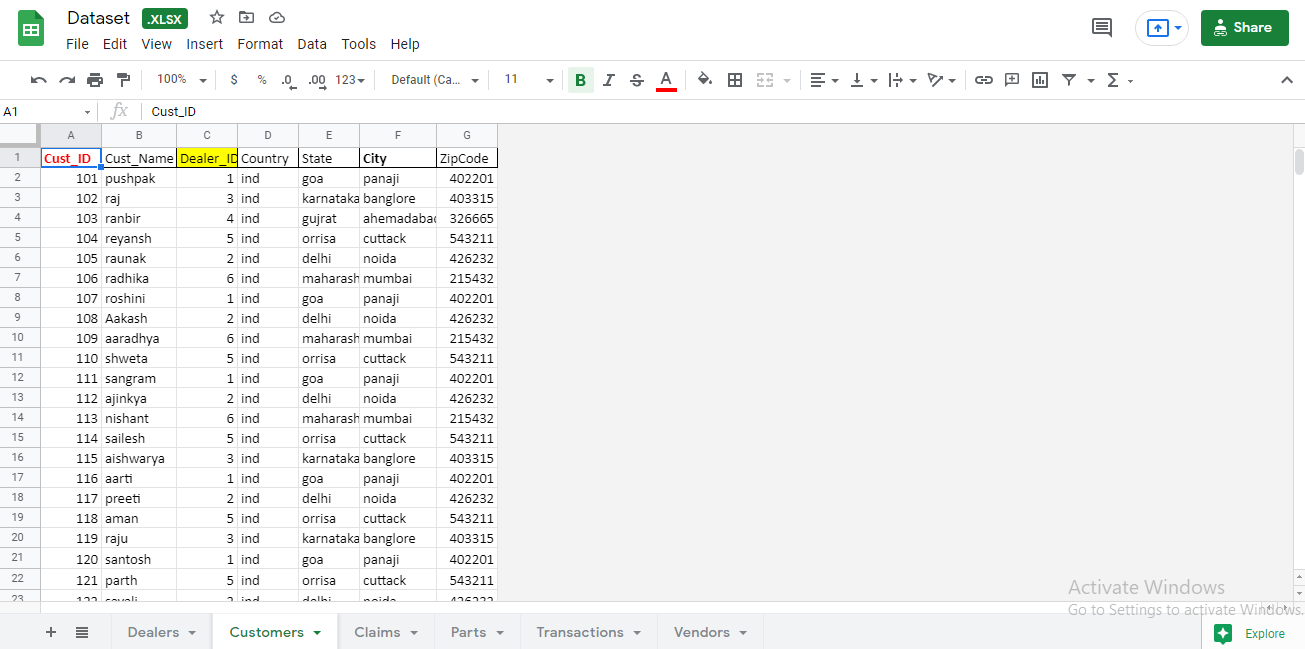
Phase 4—Model building: In this phase, you will develop datasets for training and testing purposes. Here you need to consider whether your existing tools will suffice for running the models or it will need a more robust environment (like fast and parallel processing).

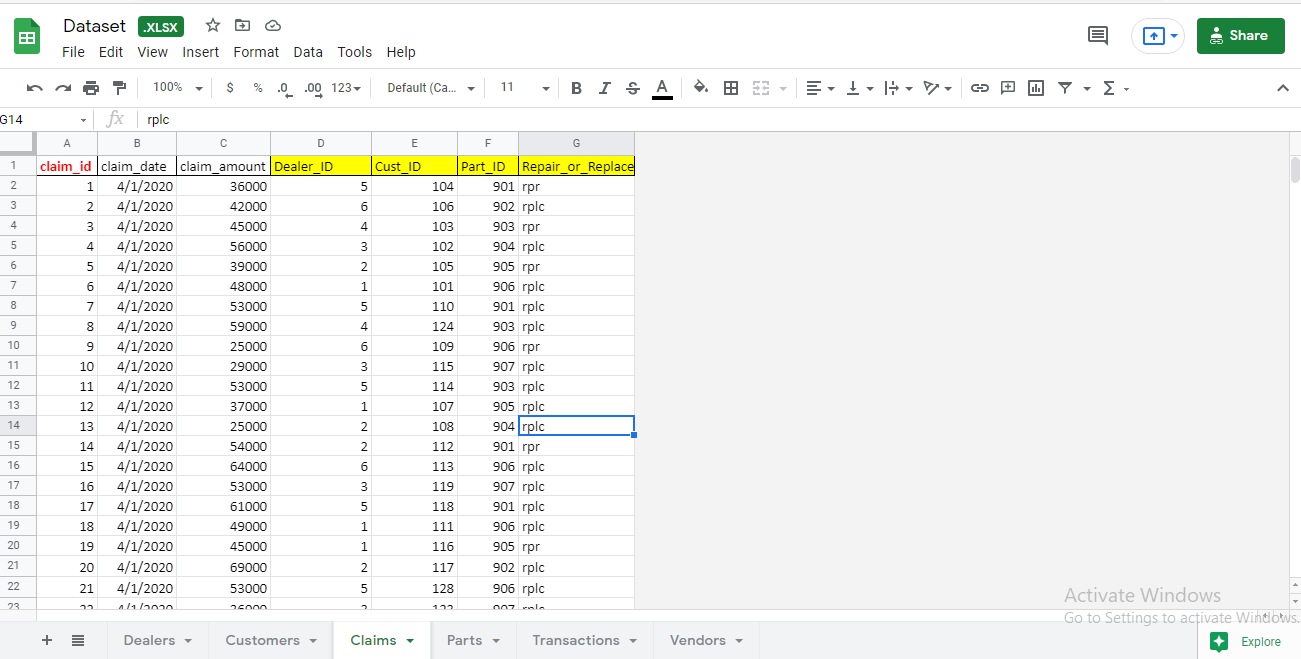
Phase 5—operationalize:   In this phase, you deliver final reports, briefings, and code and technical documents. In addition, sometimes a pilot project is also implemented in a real-time production environment. This will provide you a clear picture of the performance and other related constraints on a small scale before full deployment.

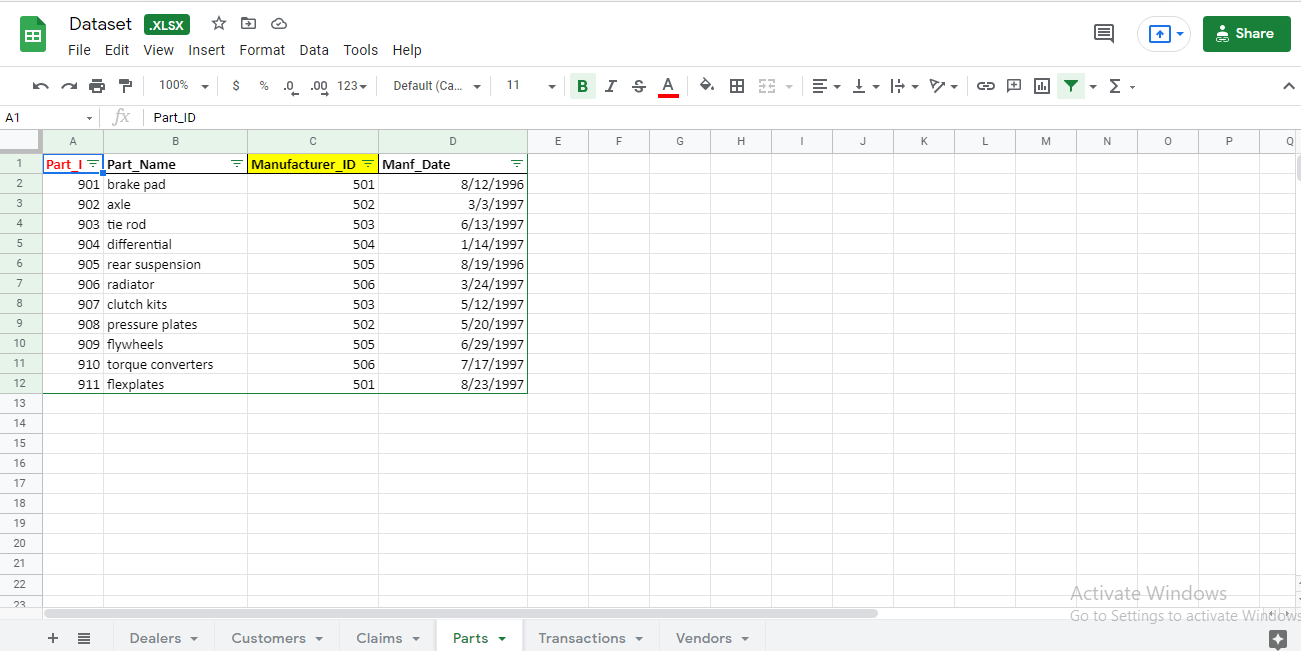
**Phase 6—Communicate results:**Now it is important to evaluate if you have been able to achieve your goal that you had planned in the first phase. So, in the last phase, you identify all the key findings, communicate to the stakeholders and determine if the results of the project are a success or a failure based on the criteria developed in Phase.

**Datasets**

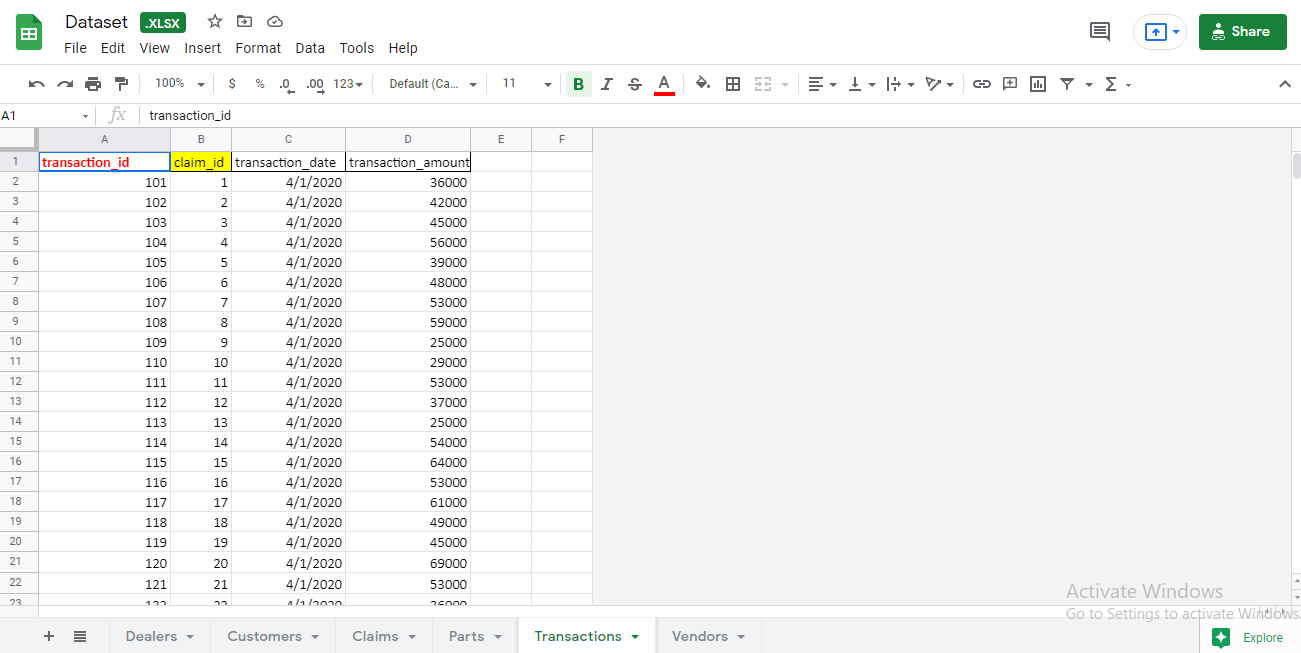
Dealers:- 

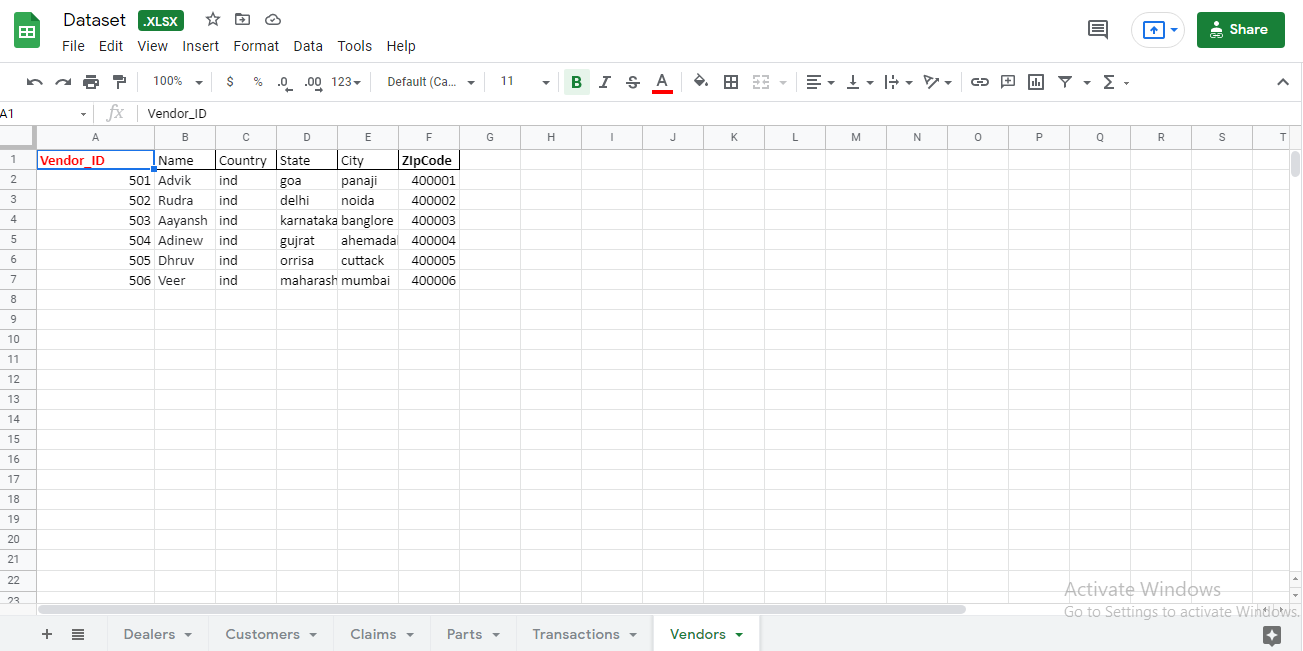
Customers:- 

Claims:- 

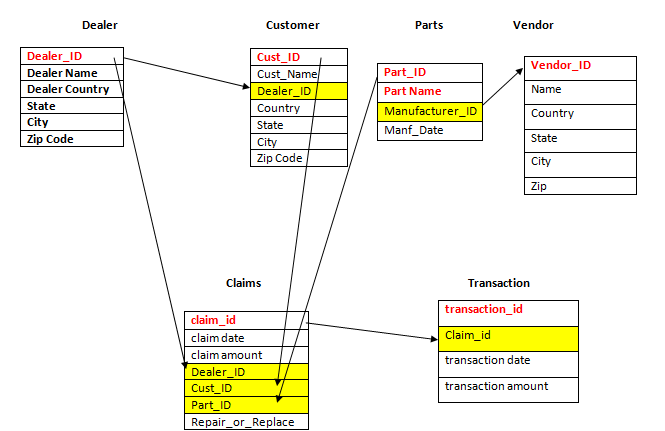
Parts:- 

Transactions:-



Vendors:- 

**Class-diagram**



**Source Code**

**import numpy as np**

**import pandas as pd**

**import matplotlib.pyplot as plt**

**Import seaborn as sns**

**from google.colab import files**

**Data\_files = files.upload()**

**dealer\_df = pd.read\_excel("Dataset.xlsx", sheet\_name = "Dealers")**

**Customers\_df = pd.read\_excel("Dataset.xlsx", sheet\_name = "Customers")**

**Claims\_df = pd.read\_excel("Dataset.xlsx", sheet\_name = "Claims")**

**Parts\_df = pd.read\_excel("Dataset.xlsx", sheet\_name = "Parts")**

**Transactions\_df = pd.read\_excel("Dataset.xlsx", sheet\_name = "Transactions")**

**Vendor\_df = pd.read\_excel("Dataset.xlsx", sheet\_name = "Vendors")**

**Parts\_df.head()**

**Claims\_df.head()**

**Customers\_df.head()**

**dealer\_df.head()**

**x = Claims\_df["Dealer\_ID"].value\_counts()**

**x.head(10)**

**Claims\_df["Part\_ID"].value\_counts().head(10)**

**Claims\_df.head()**

**Parts\_df.head()**

**Transactions\_df.head()**

**Vendor\_df.head()**

**mergedclaim\_transaction = pd.merge(Claims\_df,Transactions\_df, how = 'left',on ='claim\_id')**

**mergedclaim\_transaction.head()**

**sns.barplot(x="Part\_ID",y = 'transaction\_amount',hue ='Repair\_or\_Replace',data = mergedclaim\_transaction)**

**mergedclaim\_transaction["Vendor\_Location"] = mergedclaim\_transaction['Dealer\_ID'].map(dealer\_df['City'])**

**mergedclaim\_transaction.head()**

**sns.barplot(y='transaction\_amount',x = 'Vendor\_Location',hue = 'Repair\_or\_Replace',data = mergedclaim\_transaction)**

**mergedclaim\_transaction.groupby("City)['Part\_ID'].value\_counts().head()**

**mergedclaim\_transaction = pd.merge(mergedclaim\_transaction,Parts\_df,how ='left',on ='Part\_ID')**

**mergedclaim\_transaction['Age'] = mergedclaim\_transaction['claim\_date'] - mergedclaim\_transaction['Manf\_Date']**

**mergedclaim\_transaction**

**import datetime**

**mergedclaim\_transaction["Age"] = mergedclaim\_transaction["Age"].apply(lambda x: x.days)**

**sns.violinplot(y='Age',x = 'Part\_ID',hue = 'Repair\_or\_Replace',data = mergedclaim\_transaction)**

**sns.jointplot(y='Age',x = 'Part\_ID',hue = 'Repair\_or\_Replace',data = mergedclaim\_transaction)**

**mergedclaim\_transaction.columns**

**'Vendor\_Location\_Naoida','Vendor\_Location\_Panji']]**

**y=mergedclaim\_transaction['Repair\_or\_Replace\_rpr']**

**from sklearn.tree import DecisionTreeClassifier**

**classifier = DecisionTreeClassifier()**

**cross\_val\_score(classifier,x1,y1,cv=10,scoring='accuracy').mean()**

**from sklearn.ensemble import RandomForestClassifier**

**rfc = RandomForestClassifier(n\_estimators = 100)**

**cross\_val\_score(rfc,x1,y1,cv=10,scoring='accuracy').mean()**

**from sklearn.tree import DecisionTreeClassifier**

**classifier = DecisionTreeClassifier()**

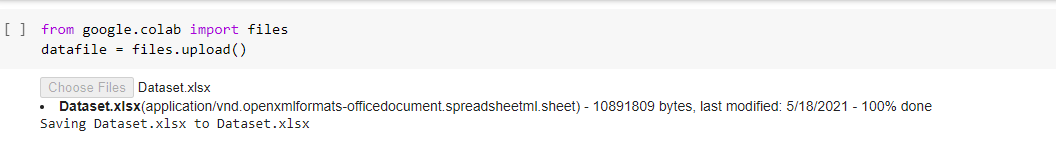
**cross\_val\_score(classifier,x2,y2,cv=10,scoring='accuracy').mean()**

**OUTPUT**

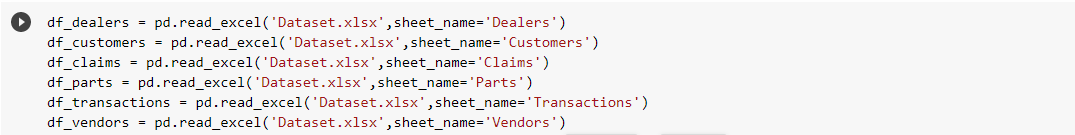
**Importing libraries**



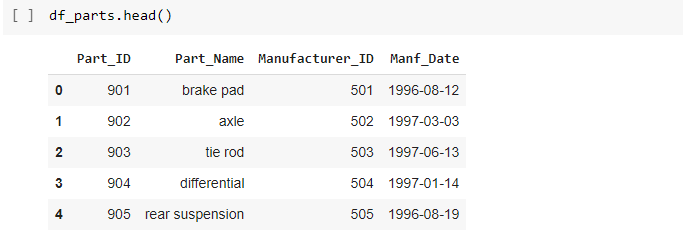
**Loading Data**

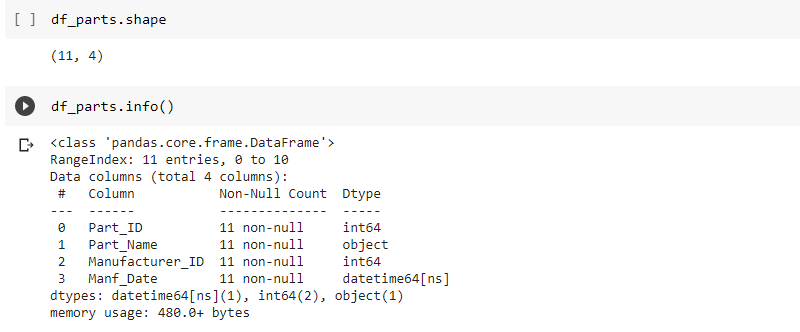


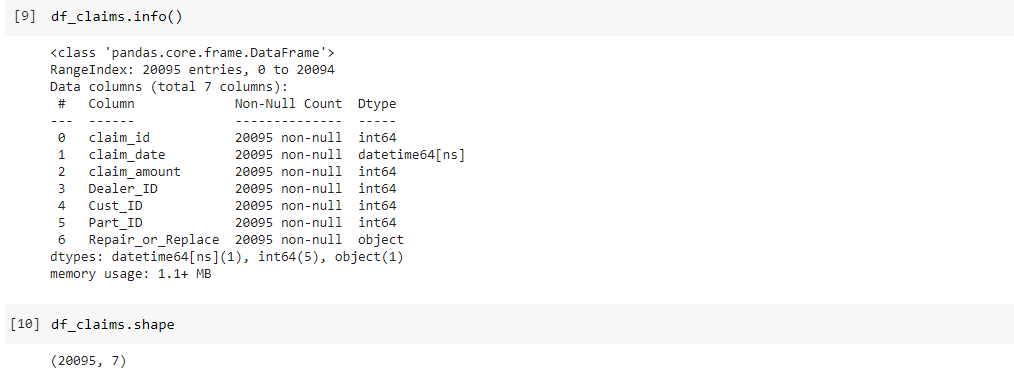
**Reading Data**

****

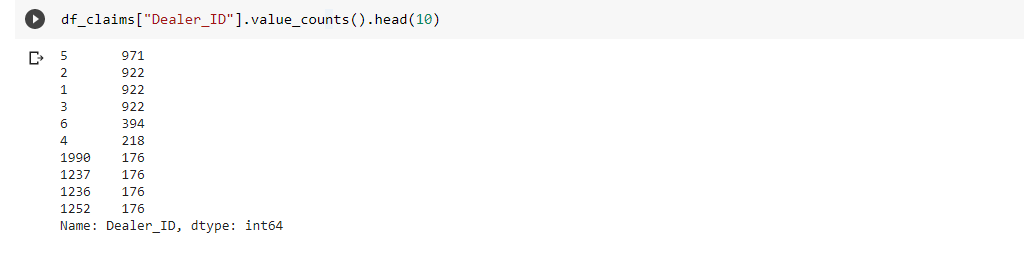
**Data Preparations  and Cleaning**

****

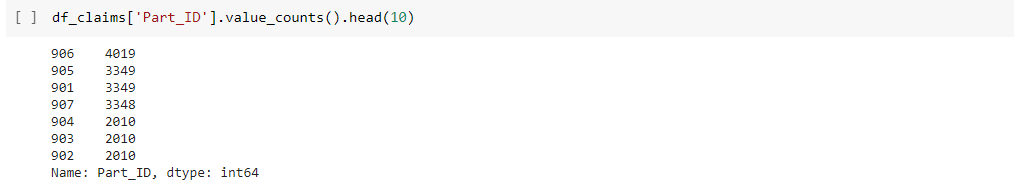


****

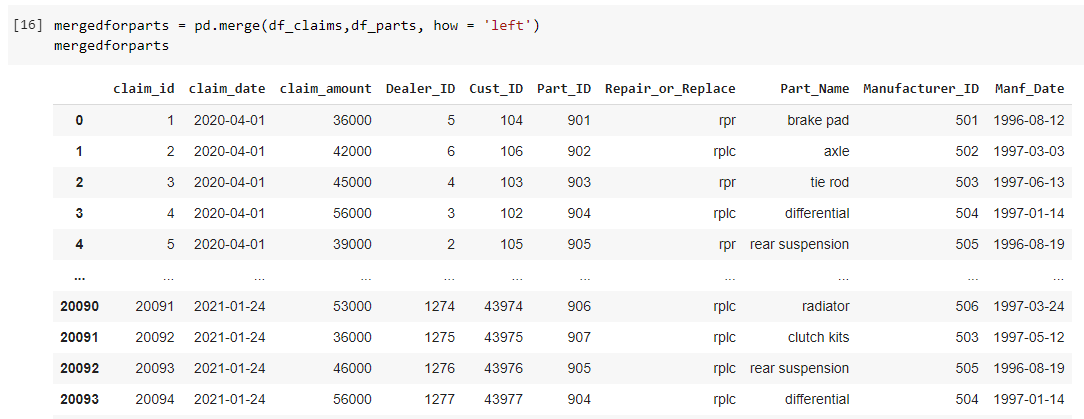
**Monthly report of top  Dealers who have claimed with high count**

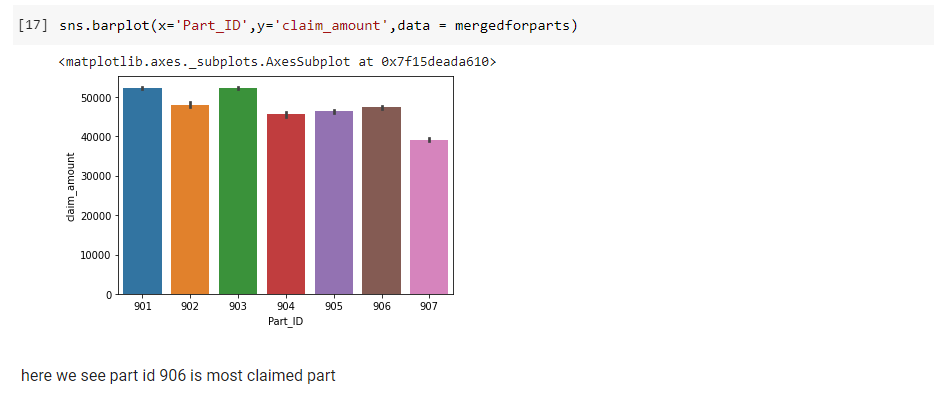


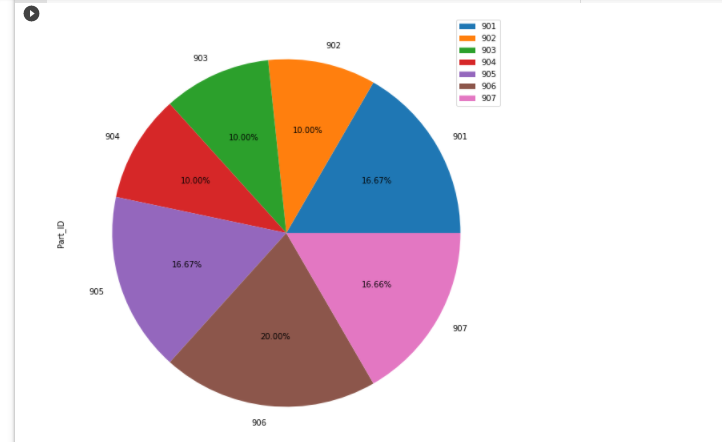
**Monthly report of top  parts  which have high failed count**



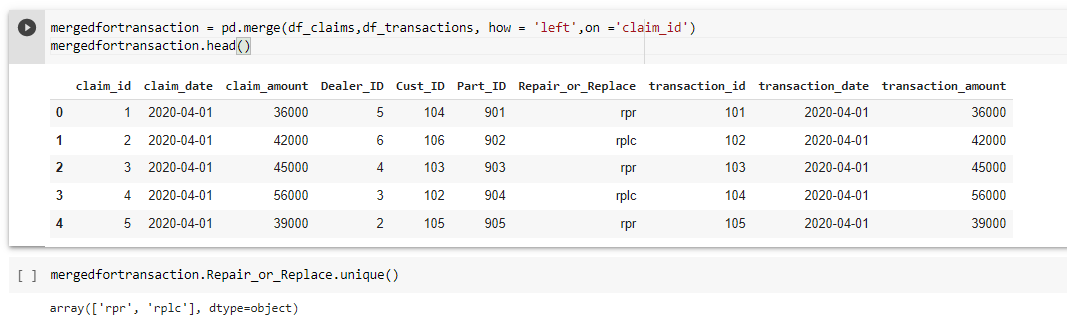
# Exploratory Data Analysis (EDA)

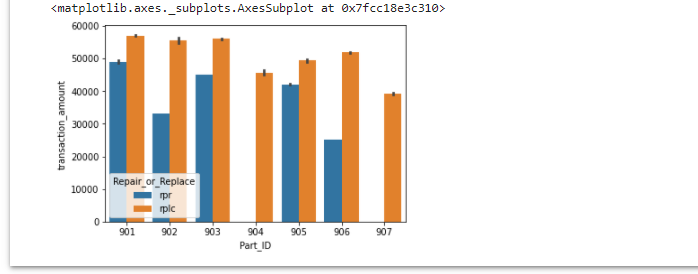




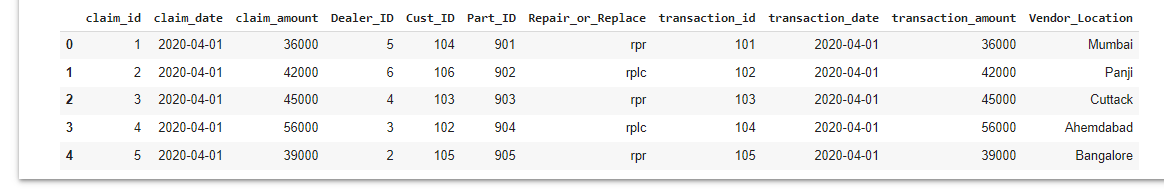


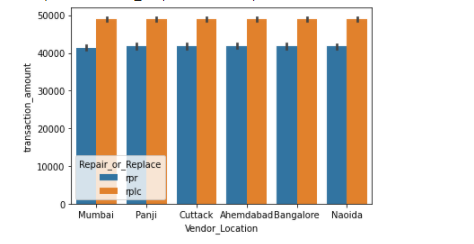
# ****Transaction amount of repair and replace by part****

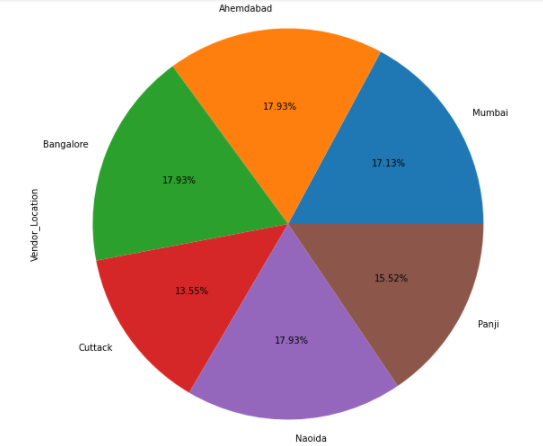




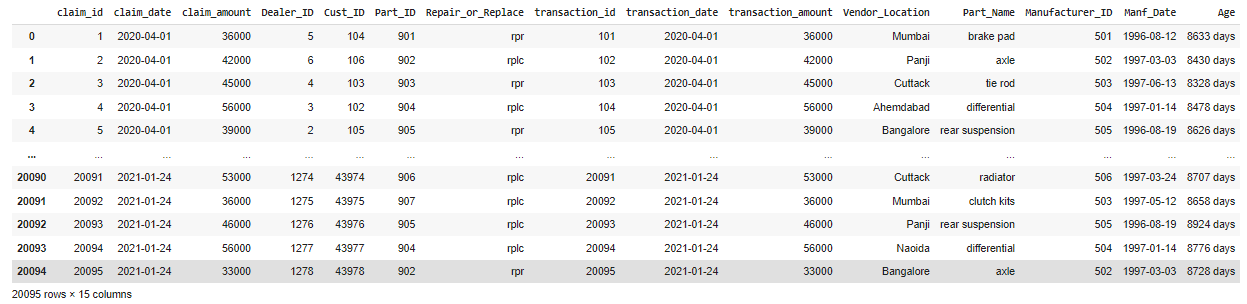
# ****Predicting the parts that may fail according to the location****

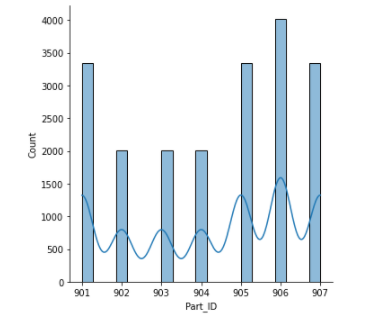


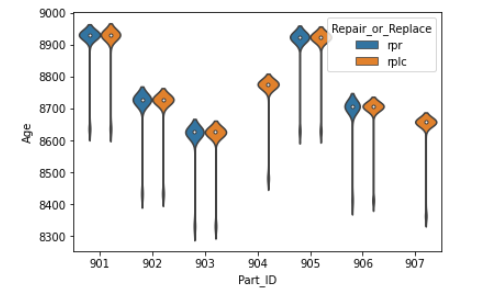


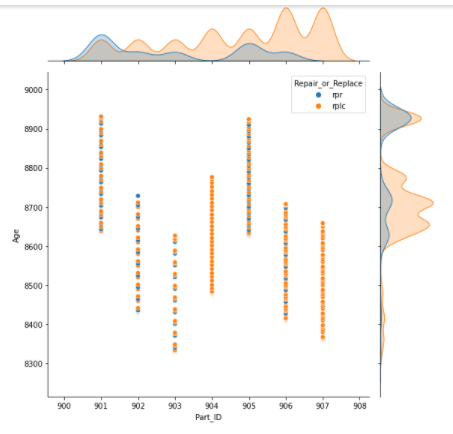


# ****Predicting the parts that may fail according to the Age****

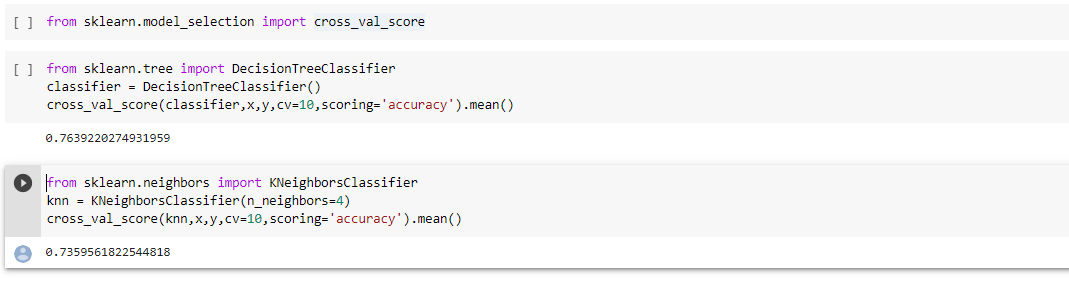








# ****Predicting Part Failure****



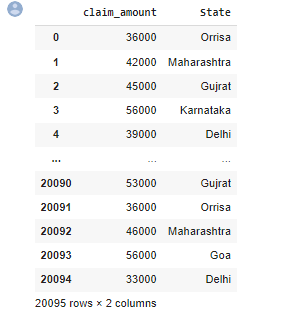
here we see best accuracy by DecisionTreeCLassifier ,so we are using decision tree classifier for further process

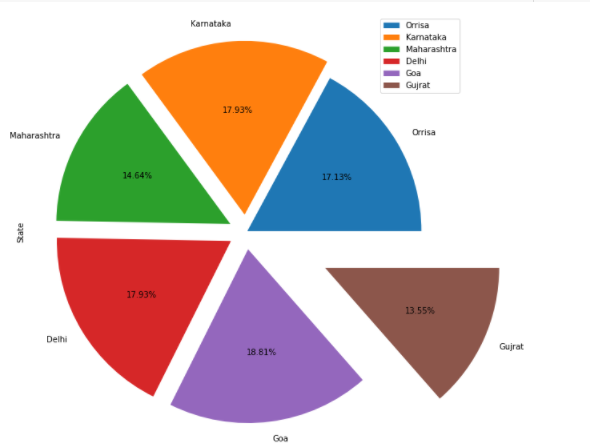
# ****To determine whether the defective part is to be replaced or repaired****

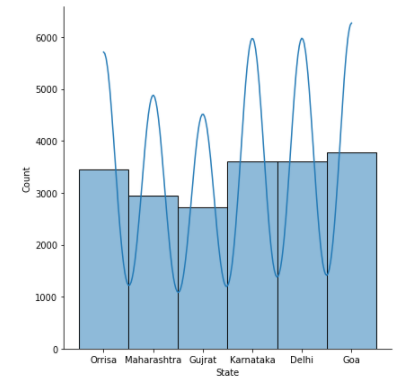
# To determine whether the defective part is to be replaced or repaired.png

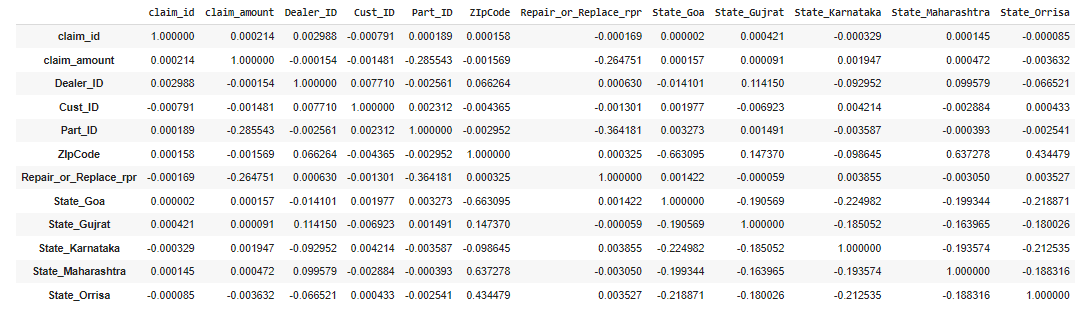
# To determine whether the defective part is to be replaced or repaired2.png

# ****Processing of variety of data****











**Conclusion.**

* **Dealer ID 5  claimed parts monthly 971 times.**
* **Part\_ID  906 is most failed part.**
* **Majority of the Parts are being replaced.**
* **Parts in Goa are among the most claimed parts**.

**Silent Features**

* **Easy to understand :-**

The structure of the system makes it easy to understand and operate the whole system.

* **User Friendly -:**

The system is user friendly and data can be backup and recovered easily**.**

* **System Reliability-:**

The system is easy to understand and designed in a user-friendly nature. It takes the less amount of time to understand the basic concept of operation .

* **System Accuracy -:**

The system is capable for doing correct calculations and generates correct reports .

* **Time saving -:**

Our system saves more time than the Manual system.

**Bibliography.**

# Books -: Python Programming Fundamentals- A Beginner's Handbook by  [Nischay kumar Hegde](https://www.amazon.in/s/ref=dp_byline_sr_book_1?ie=UTF8&field-author=Nischay+kumar+Hegde&search-alias=stripbooks)

**References for the Project Development were taken from the following Web Sites.**

**Websites -:**

<http://www.wikipedia.org/>

<https://www.kaggle.com/>

<https://colab.research.google.com/notebooks/intro.ipynb>

<https://pandas.pydata.org/>

<https://seaborn.pydata.org/>

<https://scikit-learn.org/stable/>