****

**OUTPATEINT PAYMENT OF A HOSPITAL**

Submitted

By

Ankit Yadav

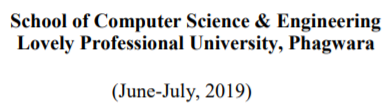
11707994

CSE B.TECH MBA

Under the Guidance of

**Amith Ashokan**

Industry Coordinator



**DECLARATION**

I hereby declare that I have completed my six weeks summer training at Splen Technologies, Lucknow from 7 June to 10 July under the guidance of Amith Ashokan. I have declare that I have worked with full dedication during these six weeks of training and my learning outcomes fulfill the requirements of training for the award of degree of CSE B.TECH MBA, Lovely Professional University, Phagwara.

(Signature of student)

Ankit Yadav

11707994

Date:\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_\_

**Index :**

|  |  |  |
| --- | --- | --- |
| **SR NO:** | **TOPIC:** | **PAGE NO:** |
| 1: | INTRODUCTION |  |
| 2: | PROFILE OF THE PROBLEM |  |
| 3: | SOFTWARE REQUIREMENT ANALYSIS |  |
| 4: | TESTING & IMPLEMENTATION |  |
| 5: | GANTT CHART |  |
| 6: | PROJECT LEGACY |  |
| 7: | BIBLIOGRAPHY |  |

**1.INTRODUCTION**

**1.1 Data:**

[Raw data](https://en.wikipedia.org/wiki/Raw_data) ("unprocessed data") is a collection of [numbers](https://en.wikipedia.org/wiki/Number) or [characters](https://en.wikipedia.org/wiki/Character_(computing)) before it has been "cleaned “and corrected by researchers. Raw data needs to be corrected to remove [outliers](https://en.wikipedia.org/wiki/Outlier) or obvious instrument or data entry errors (e.g., a thermometer reading from an outdoor Arctic location recording a tropical temperature).

Data is [measured](https://en.wikipedia.org/wiki/Measurement), [collected and reported](https://en.wikipedia.org/wiki/Data_reporting), and [analyzed](https://en.wikipedia.org/wiki/Data_analysis), whereupon it can be [visualized](https://en.wikipedia.org/wiki/Data_visualization) using graphs, images or other analysis tools. Data as a general [concept](https://en.wikipedia.org/wiki/Concept) refers to the fact that some existing [information](https://en.wikipedia.org/wiki/Information) or [knowledge](https://en.wikipedia.org/wiki/Knowledge) is [*represented*](https://en.wikipedia.org/wiki/Knowledge_representation_and_reasoning) or [*coded*](https://en.wikipedia.org/wiki/Code) in some form suitable for better usage or [processing](https://en.wikipedia.org/wiki/Data_processing).

[Knowledge](https://en.wikipedia.org/wiki/Knowledge) is the understanding based on extensive experience dealing with information on a subject. For example, the height of [Mount Everest](https://en.wikipedia.org/wiki/Mount_Everest) is generally considered data. The height can be measured precisely with an [altimeter](https://en.wikipedia.org/wiki/Altimeter) and entered into a database.

**1.2 Data Collection:**

Gathering data can be accomplished through a primary source (the researcher is the first person to obtain the data) or a secondary source (the researcher obtains the data that has already been collected by

other sources, such as data disseminated in a scientific journal). Data analysis methodologies vary and include data triangulation and data percolation.

The data is thereafter "percolated" using a series of pre-determined steps so as to extract the most relevant information. Although data is also increasingly used in other fields, it has been suggested that the highly interpretive nature of them might be at odds with the ethos of data as "given".

1.3: Distribution of Data

**1.4 Database:**

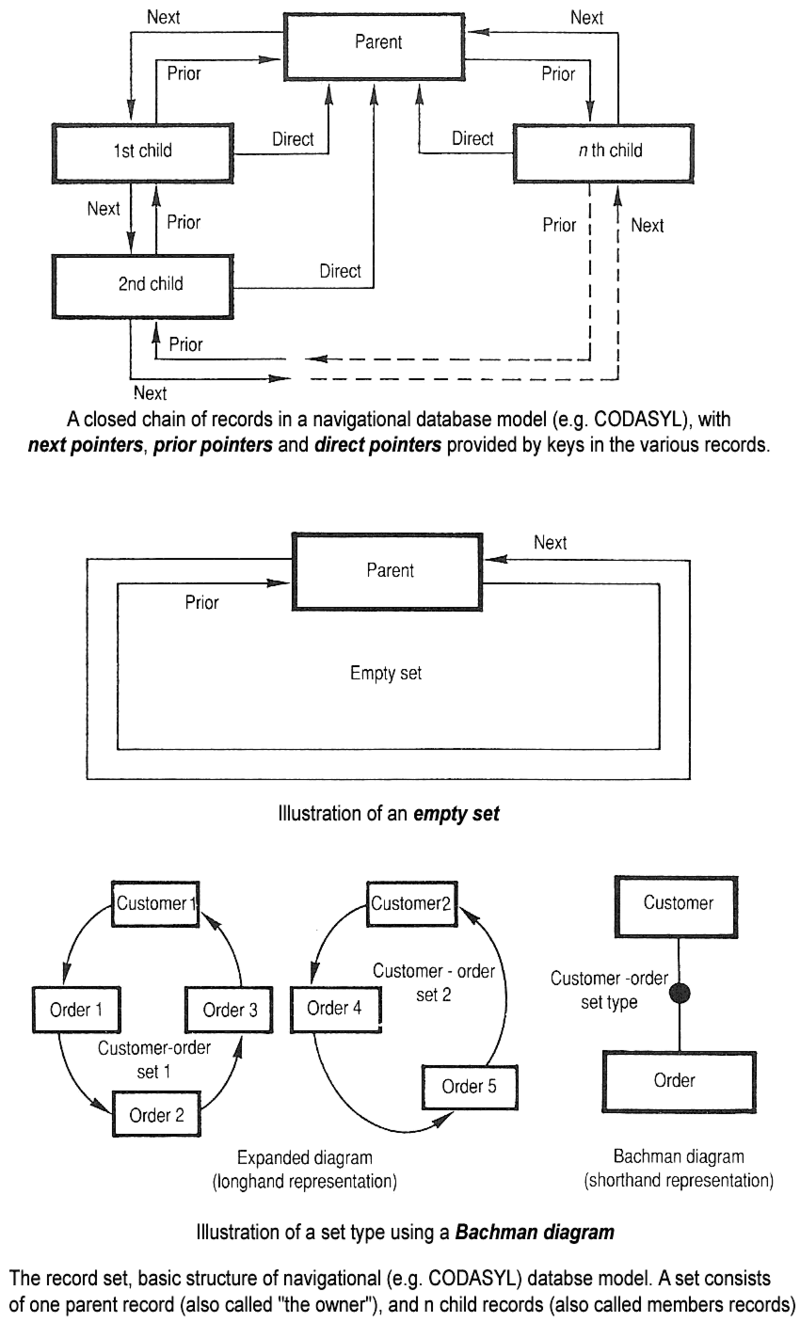
A **database** is an organized collection of [data](https://en.wikipedia.org/wiki/Data_(computing)), generally stored and accessed electronically from a computer system. Where databases are more complex they are often developed using formal [design and modeling](https://en.wikipedia.org/wiki/Database#Design_and_modeling) techniques.

The [database management system](https://en.wikipedia.org/wiki/Database#Database_management_system) (DBMS) is the [software](https://en.wikipedia.org/wiki/Software) that interacts with [end users](https://en.wikipedia.org/wiki/End_user), applications, and the database itself to capture and analyze the data. The DBMS software additionally encompasses the core facilities provided to administer the database. The sum total of the database, the DBMS and the associated applications can be referred to as a "database system". Often the term "database" is also used to loosely refer to any of the DBMS, the database system or an application associated with the database.

Existing DBMSs provide various functions that allow management of a database and its data which can be classified into four main functional groups:

* **Data definition** – Creation, modification and removal of definitions that define the organization of the data.
* **Update** – Insertion, modification, and deletion of the actual data.
* **Retrieval** – Providing information in a form directly usable or for further processing by other

applications. The retrieved data may be made available in a form basically the same as it is stored in the database or in a new form obtained by altering or combining existing data from the database.



1.5 : Flow Diagram of Database

**1.6: SUPERVISED LEARNING**

**Supervised learning** is the [machine learning](https://en.wikipedia.org/wiki/Machine_learning) task of learning a function that maps an input to an output based on example input-output pairs.It infers a function from labeled [training data](https://en.wikipedia.org/wiki/Training_set) consisting of a set of training examples. In supervised learning, each example is a pair consisting of an input object (typically a vector) and a desired output value (also called the supervisory signal). A supervised learning algorithm analyzes the training data and produces an inferred function, which can be used for mapping new examples.

**1.7 UNSUPERVISED LEARNING**

**Unsupervised learning** is a type of self-organized [Hebbian learning](https://en.wikipedia.org/wiki/Hebbian_learning" \o "Hebbian learning) that helps find previously unknown patterns in data set without pre-existing labels. It is also known as [self-organization](https://en.wikipedia.org/wiki/Self-organization) and allows modeling [probability densities](https://en.wikipedia.org/wiki/Probability_density_function) of given inputs.[[1]](https://en.wikipedia.org/wiki/Unsupervised_learning#cite_note-Hinton99a-1) It is one of the main three categories of machine learning, along with [supervised](https://en.wikipedia.org/wiki/Supervised_learning) and [reinforcement learning](https://en.wikipedia.org/wiki/Reinforcement_learning). Semi-supervised learning has also been described, and is a hybridization of supervised and unsupervised techniques.

**1.8 REGRESSION TECHNIQUES**

Linear and Logistic regressions are usually the first modelling algorithms that people learn for Machine Learning and Data Science. Both are great since they’re easy to use and interpret. However, their inherent simplicity also comes with a few drawbacks and in many cases they’re not really the best choice of regression model. There are in fact several different types of regressions, each with their own strengths and weaknesses.

## Linear Regression

## Lasso Regression

## Elastic Net Regression

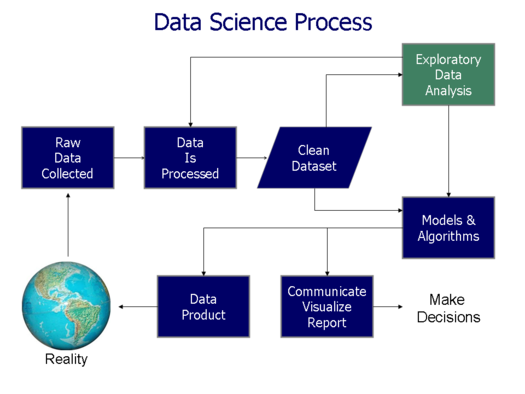
Polynomial Regression

Ridge Regression

**1.9 DATA VISUALISATION**

Data visualization refers to the techniques used to communicate data or information by encoding it as visual objects (e.g., points, lines or bars) contained in graphics. The goal is to communicate information clearly and efficiently to users. It is one of the steps in [data analysis](https://en.wikipedia.org/wiki/Data_analysis) or [data science](https://en.wikipedia.org/wiki/Data_science). According to Friedman (2008) the "main goal of data visualization is to communicate information clearly and effectively through graphical means. It doesn't mean

that data visualization needs to look boring to be functional or extremely sophisticated to look beautiful. To convey ideas effectively, both aesthetic form and functionality need to go hand in hand, providing insights into a rather sparse and complex data set by communicating its key-aspects in a more intuitive way.



1.10 VISUALISATION MAP

There are various method by which we can perform Data Visualizations like using Libraries like PANDAS and MATHPLOTLIB etc.

**1.11 DIFFERENT LIBRARIES**

NumPy

## NumPy is the first choice among developers and data scientists who are aware of the technologies which are dealing with data-oriented stuff. It is a Python package available for performing scientific computations. It is registered under the BSD license.

## Theano

Theano is another useful Python library assists data scientists in performing large multi-dimensional arrays related computing operations. It is more like TensorFlow but the only difference is, it is not that efficient.

## Keras

Keras is one of the most powerful Python libraries which allow high-level neural networks APIs for integration. Theses APIs execute over the top of TensorFlow, Theano and CNTK. Keras was created for reducing challenges faced in complex researches allowing them to compute faster

## SciPy

## SciPy is another Python library for researchers, developers and data scientists. Do not get confused with the SciPy

## stack and library. It provides statistics, optimizations, integration and linear algebra packages for computation. It is

## Based on NumPy concept to deal with complex mathematical problems.

It provides numerical routines for optimization and integration. It inherits varieties of sub-modules to choose from. If you have just started your data science career, SciPy can be very helpful to guide you throughout the whole numerical computations thing.

## PANDAS

PANDAS referred as Python Data Analysis Library. PANDAS is another open source Python library for availing high-performance data structures and analysis tools. It is developed over the Numpy package. It contains Data Frame as its main data structure.

## PyBrain

PyBrain is another powerful modular ML library available in Python. PyBrain stands for Python Based Reinforcement Learning, Artificial Intelligence, and Neural Network Library. For entry-level data scientists, it offers flexible modules and algorithms for advanced research.

It has varieties for algorithms for evolution, neural networks, supervised and unsupervised learning.  For real-life tasks, it has emerged as the best tool which is developed across the neural network in the kernel.

## SciKit-Learn

Scikit-Learn is a simple tool for data analysis and mining-related tasks. It is open-source and licensed under the BSD. Anyone can access or reuse it in various contexts. SciKit is developed over the Numpy, Scipy, and Matplotlib. It is being used for classification, regression and clustering o manage spam, image recognition, drug response, stock pricing, customer segmentation etc. It also allows dimensionality reduction, model selection and pre-processing.

## Matplotlib

This 2D plotting library of Python is very famous among data scientists for designing varieties of figures in multiple formats which is compatible across their respected platforms. One can easily use it in their Python code, IPython shells or Jupyter notebook, application servers.  With Matplotlib, you can make histograms, plots, bar charts, scatter plots etc.

## Seaborn

Seaborn was designed to visualize the complex statistical models. It has the potential to deliver accurate graphs such as heat maps. Seaborn was created on the concept of Matplotlib and somehow it is highly dependent on that. Minor to minor data distributions can be easily visualized through this library which is why it has become familiar among data scientists and developers.

## Bokeh

Bokeh is one more visualization library for designing interactive plots. Just like the last one, it

is also developed on matplotlib. Due to the used data-driven documents (D3.js. support it presents interactive designs in the web browser.

## Plotly

Let’s talk about the Plotly which is one of the most famous web-based frameworks for data scientists. This toolbox offers designing of visualization models with varieties of APIs supported by multiple programing languages including Python. You can easily use interactive graphics and numerous robust accessible through its main website plot.ly. For using Plotly in your working model you need to set up available API keys properly.

We have mainly used very few Libraries like pandas ,mathplotlib etc in this project.

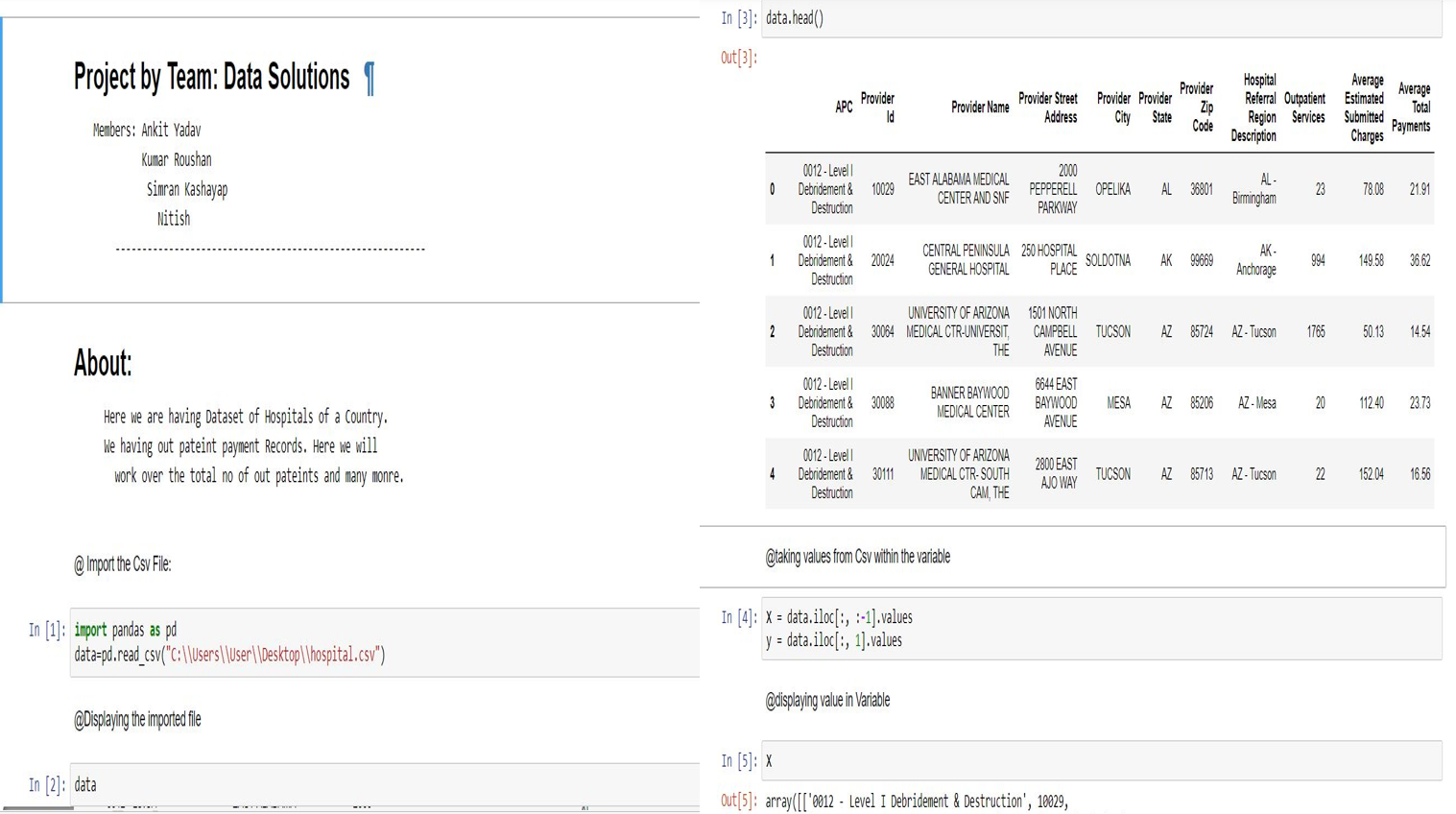
**2: PROFILE OF THE PROBLEM**

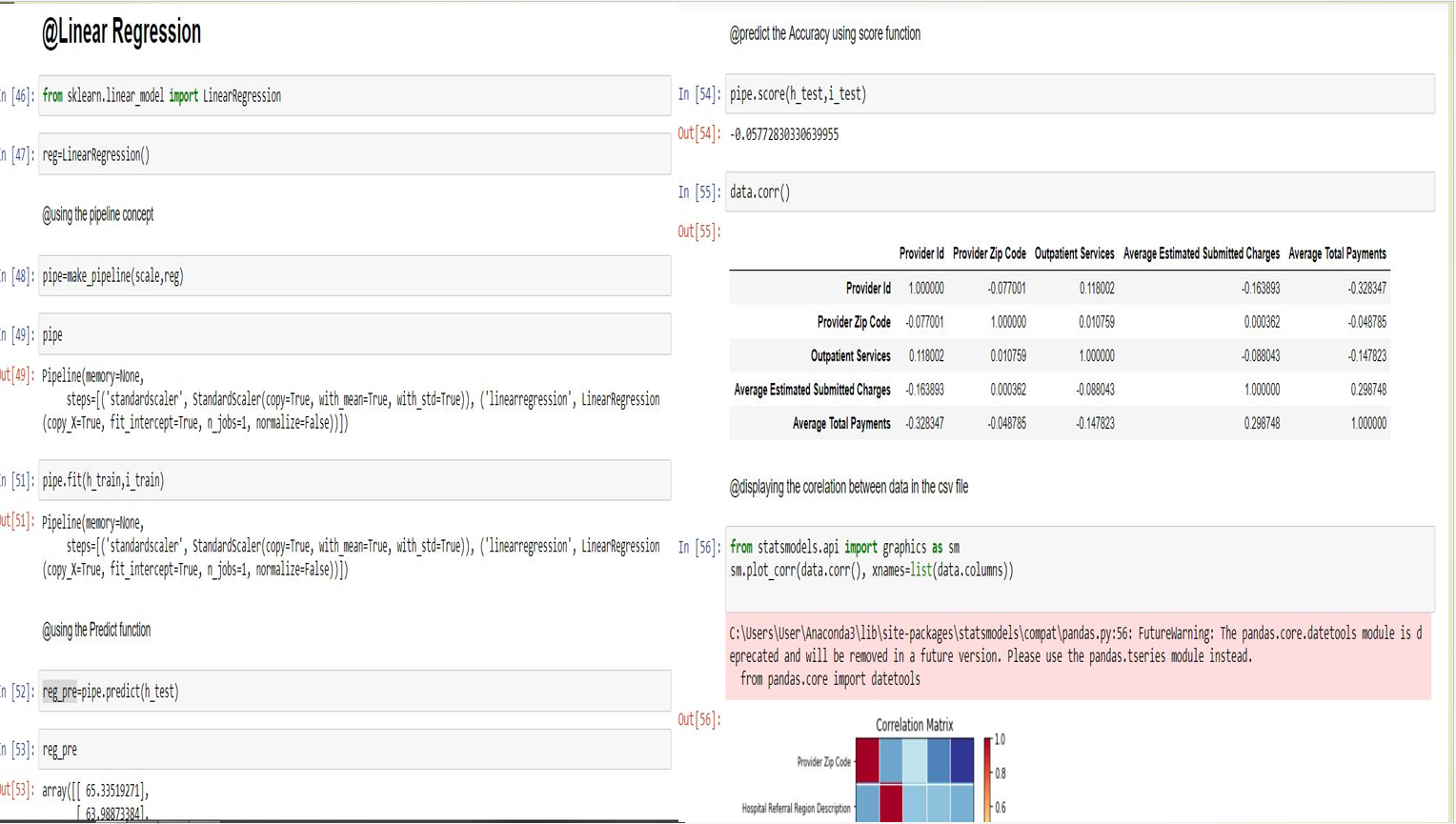
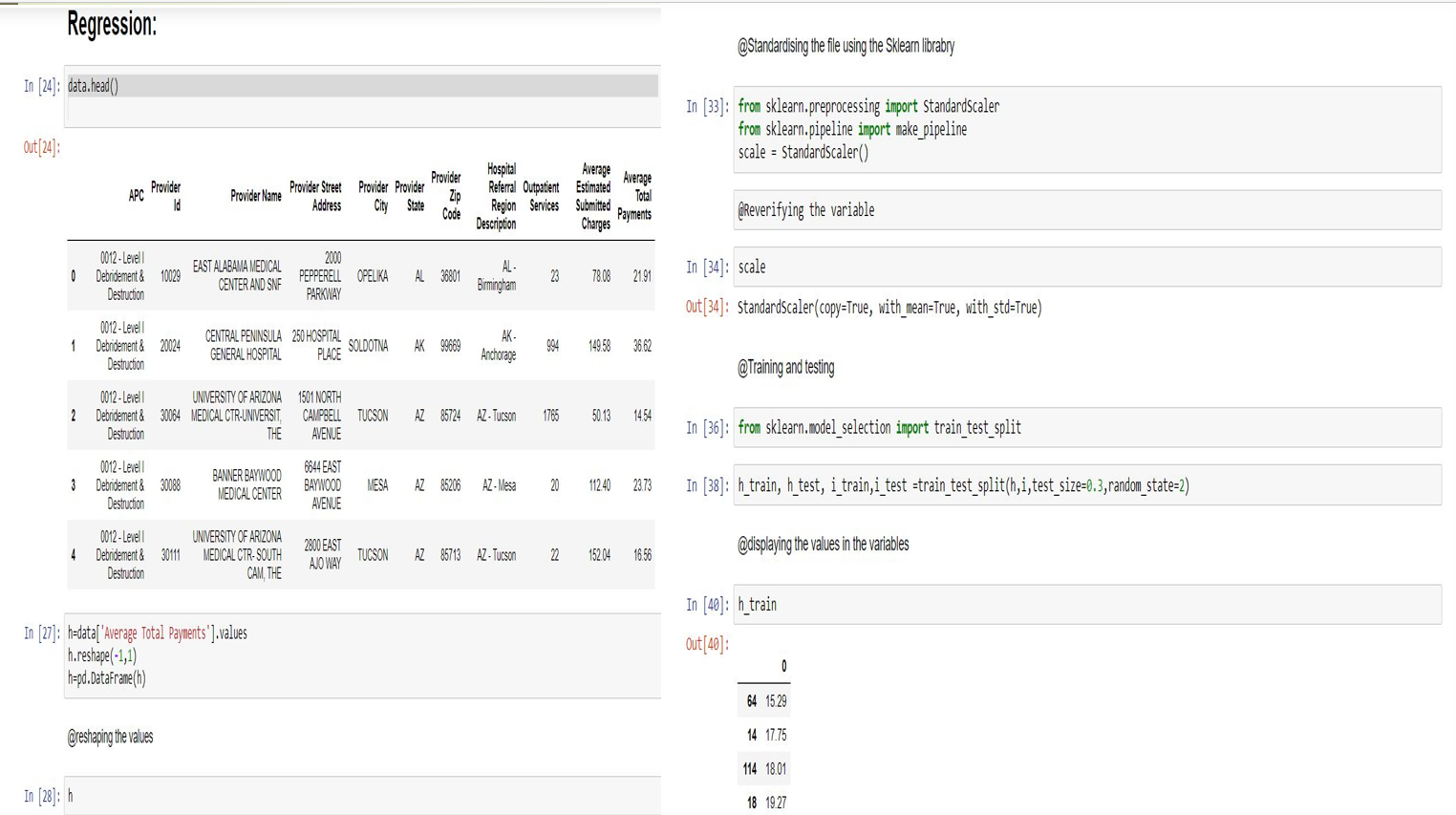
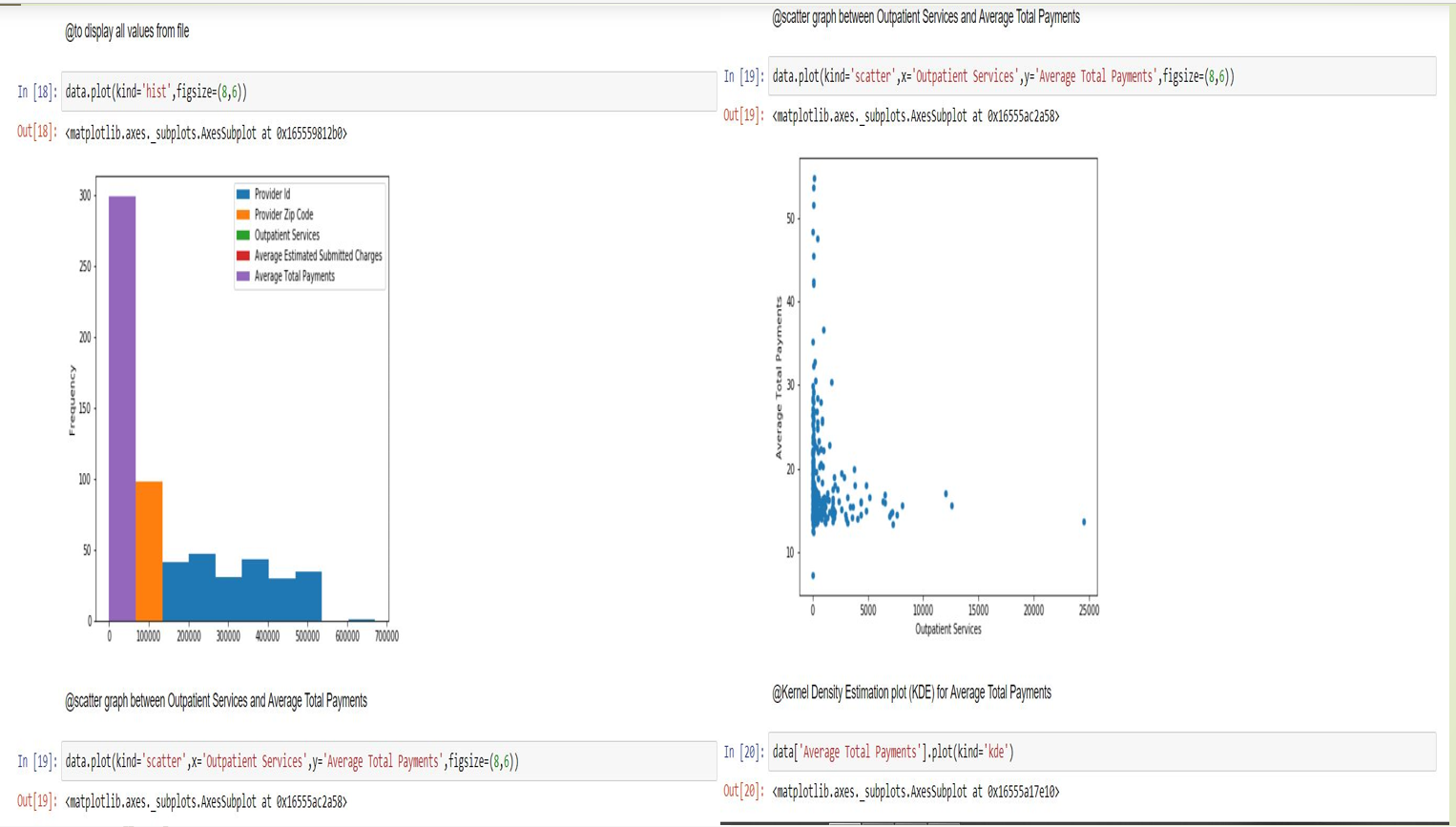
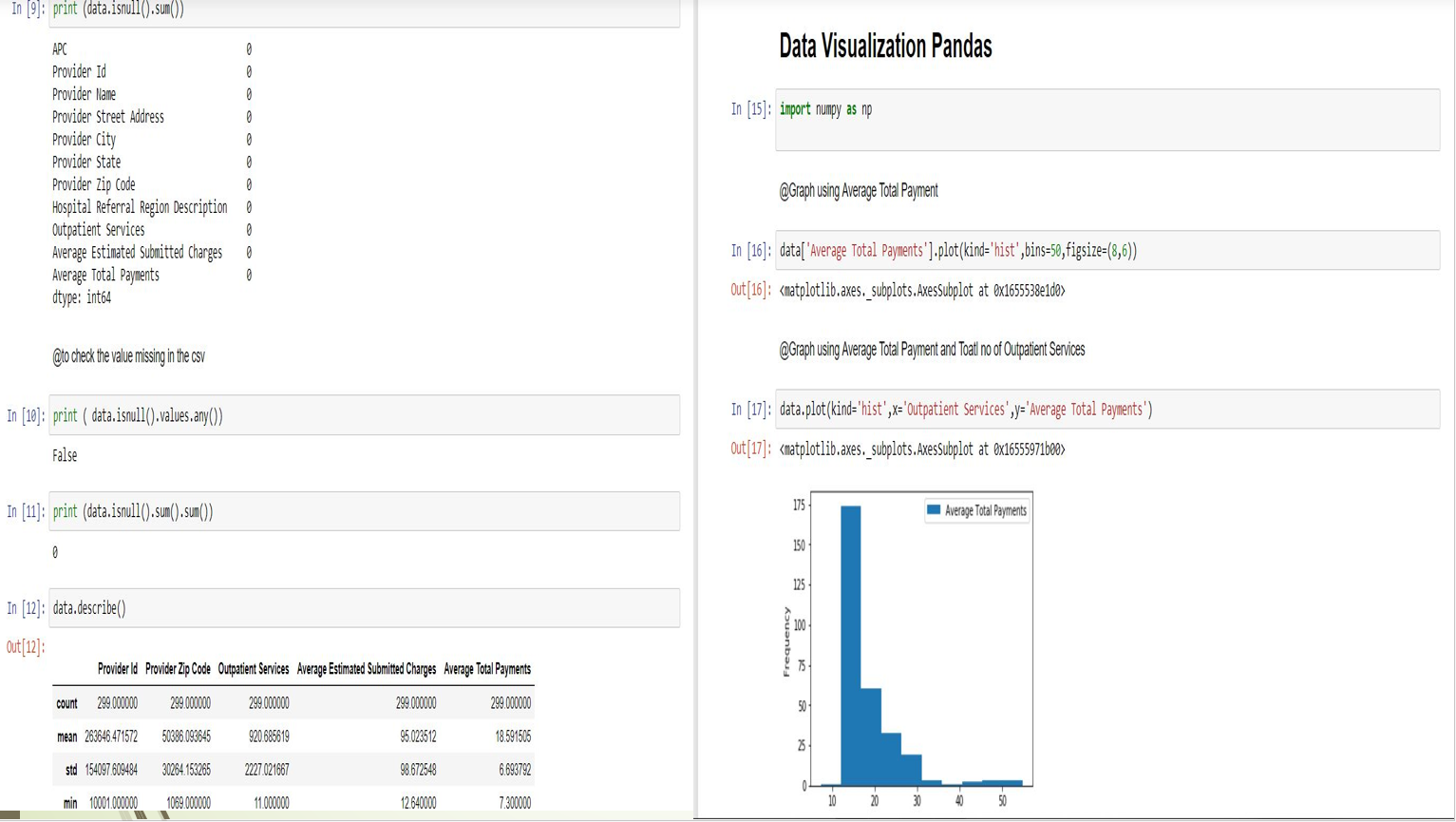
The problem is all about how to make the record of a Hospital’s Outpatient payment Record and predicate about what would be the about amount of the final outpatients for a certain amount of the Payment .Here we have display various sections by using different modules of the data science with python. Here we have to show and check whether we have tried to apply major concepts like Data cleaning, Data manipulation along with different techniques. We have to calculate the total no of Hospital who are coming the csv file many times. We have to prepare a simple and effective regression technique for the result.

**3: SOFTWARE REQUIREMENT ANALYSIS**

We require some sort of deep knowledge about python concepts and methodologies for the better implementations. We need ANACONDA Software for the project to test and implement for the result. We need Jupytor note book for writing the code and compiling it.

**4: TESTING & IMPLEMENTATION**

****

****

**5:** GANTT CHART