**PROJECT REPORT**

# **University Admission Prediction**



Submitted By: -

Name: Aishwarya Patil

USN: 3GN18CS007

Department: Computer Science

Submission Date: 13/10/2021

# **UNDERTAKING**

I declare that the work presented in this project titled “University Admission

Prediction”, submitted to the All India Council of Robotics and Automation, for the award of the Internship in Artificial Intelligence, is my original work. I have not plagiarized or submitted the same work for the award of any other Internship. In case this undertaking is found incorrect, I accept that my Project may be unconditionally withdrawn.

13 October,2021 Aishwarya Patil

3GN18CS007

# **CERTIFICATE**

This is to certify that the work contained in the project titled “University Admission Prediction”, by Aishwarya Patil, has been carried out under my supervision and that this work has not been submitted elsewhere for internship.

All India Council of Robotics and Automation

Data Sceince And Artificial Intelligence

Delhi-110020

# **ACKNOWLEDGEMENT**

I take upon this opportunity to acknowledge the many people whose prayers and support meant a lot to me.

I am deeply indebted to my mentor Mr. Sumit Chatterjee who motivated me along the way.

I would like to thank all my teachers who help me in this project.

I further thank my friends.

My heartfelt thanks to parents who supported me a lot. I owe my sincere gratitude towards the almighty God.

Finally, I would like to wind up by paying my heartfelt thanks to AICRA institute who provided me with this great opportunity.

Aishwarya Patil

**CONTENTS**

Pg.No.

1. Purpose 6
2. System Architecture 7
3. Implementation 8
4. Important Libraries 9
5. Load Dataset 10
6. Summarise the Dataset 11
7. Data Visualisation 12-14
8. Conclusion 15
9. References 16

## Purpose

To apply for a master's degree is a very expensive and intensive work. With this kernel, students will guess their capacities and they will decide whether to apply for a master's degree or not.So, basically this set is about the Graduate Admissions datai.e. Given a set of standardized scores like GRE, TOEFL, SOP standard scores, LOR standard scores, what is probability ( basically i have done a YES/NO scenario ) of gaining admission into a particular school. All those folks who are preparing for MS, might point out this question, from where did you get SOP & LOR scores. These aren’t public figures ? I mean yes, it might not be public, but don’t you think universities might be grading these applications on some scale of rating so that the scores can be standardized. Hence the SOP, LOR scores.

# **SYSTEM ARCHITECTURE**

1. The initial procedure for this project is the collection of data. We have collected the data set from Kaggle, which is available; it is an open-source.
2. The next step after data collection is data pre-processing. In this step, the data is cleansed by removing unnecessary values. It also removes the missing/ null/ corrupted values.
3. After Data is cleansed, the next step is dividing the data, we divide the data into two sets, Training data and testing data. The values must deal before we start to construct the training model. By using training data, we build a model for prediction.
4. Now, we must calculate the accuracy of the model.
5. The final step is predicting the disease.

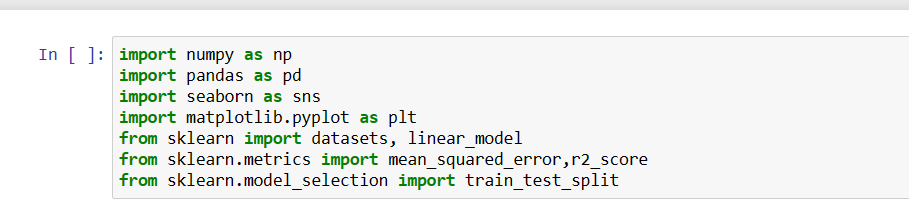
**IMPLEMENTATION**

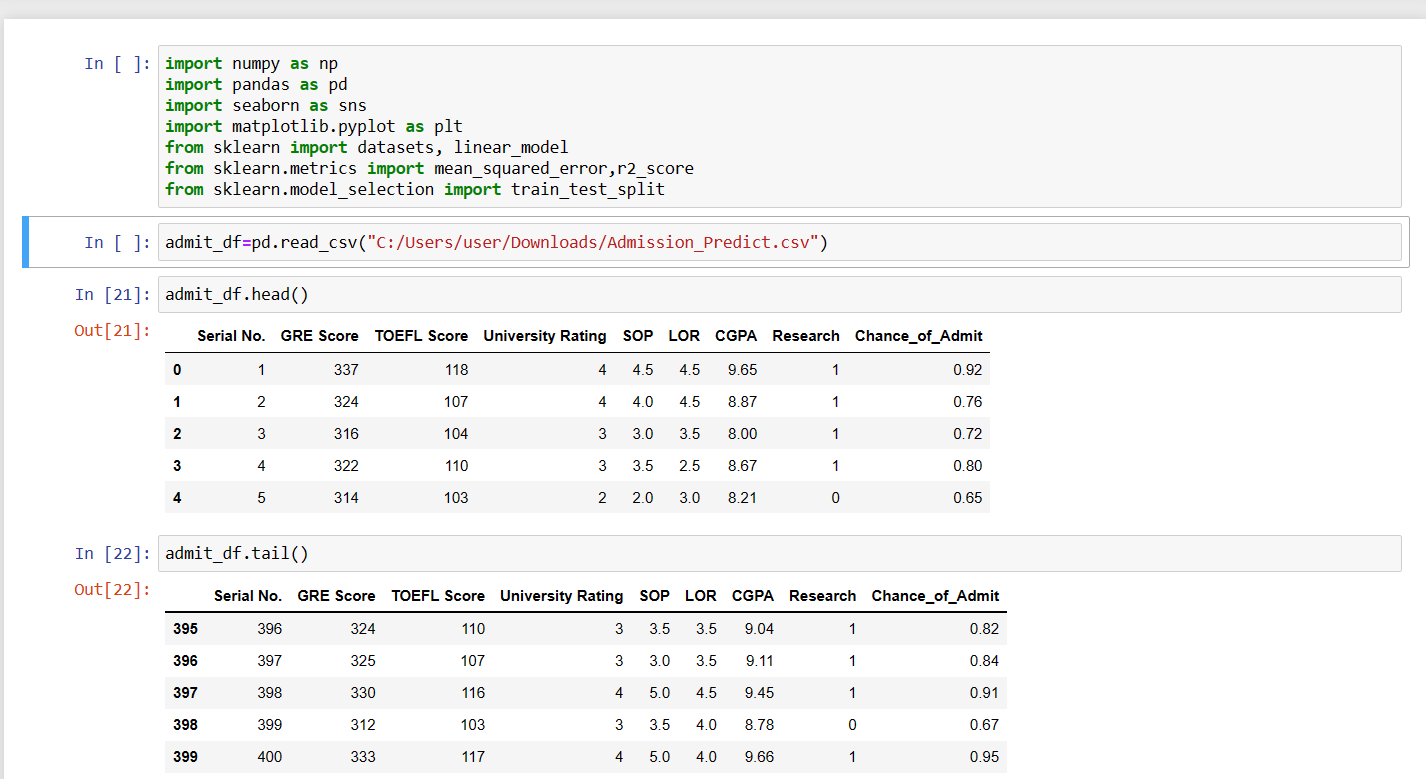
This work used Python programming for this project, as it is a high-level programming language and it has vast libraries and Python automates tasks and makes it efficient. Firstly, we need to install Python then we need to import some libraries, they are:

1. Numpy: Numpy is used for multi-dimensional arrays, it does element to element operations and it also has different methods for processing arrays.
2. Panda: Pandas is one of the highly used python libraries, it provides high performance. It manipulates data and it makes data analysis fast and easy.
3. Sklearn: It is most useful library, this library contains lot of efficient tools, it is used to build models like statistical modelling including classification, regression, clustering. After loading required packages, we divide dataset as training and testing as follows, here 80 % of dataset is taken as training and remaining 20 % as to perform test.

# **IMPORT LIBRARIES**

First, let’s import all the modules, functions and objects we are going to use

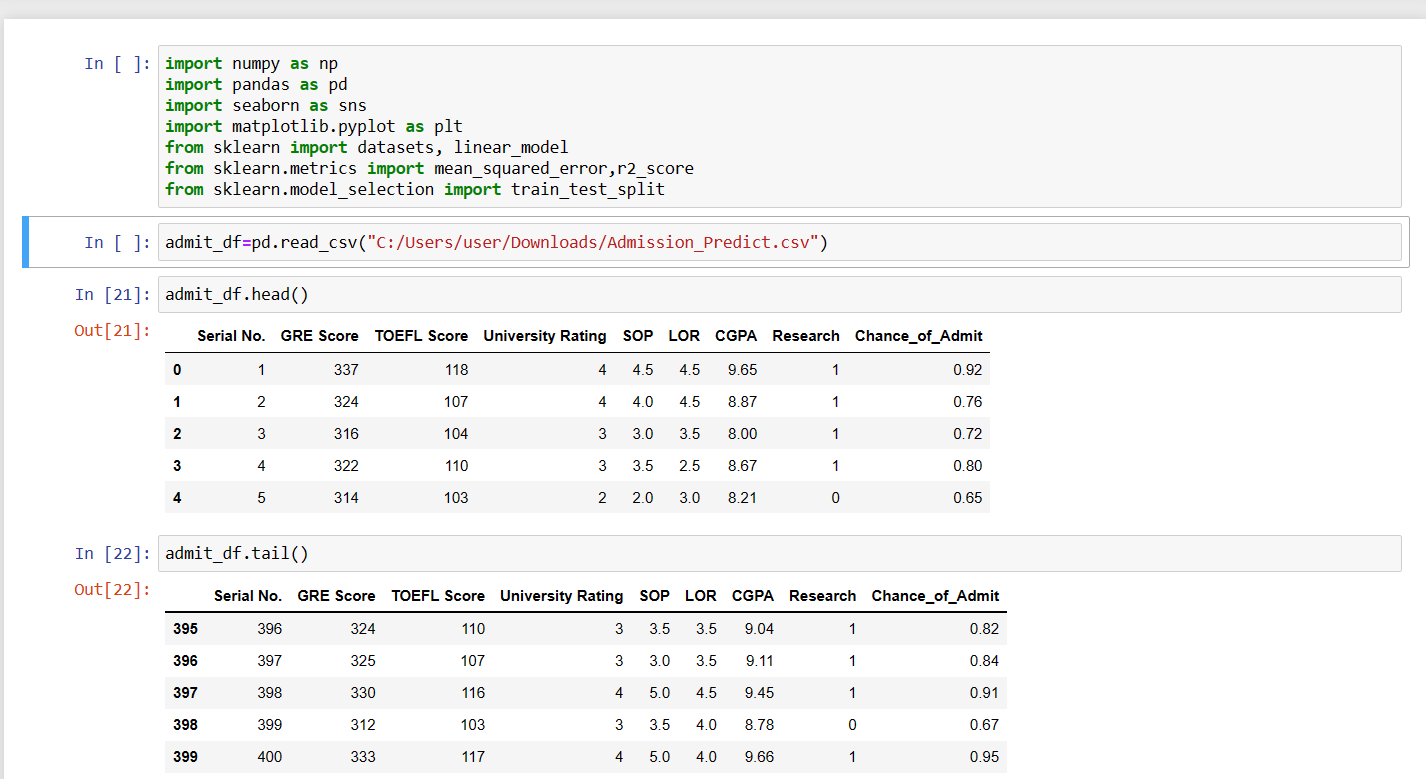




\*\*

# **Load Dataset**

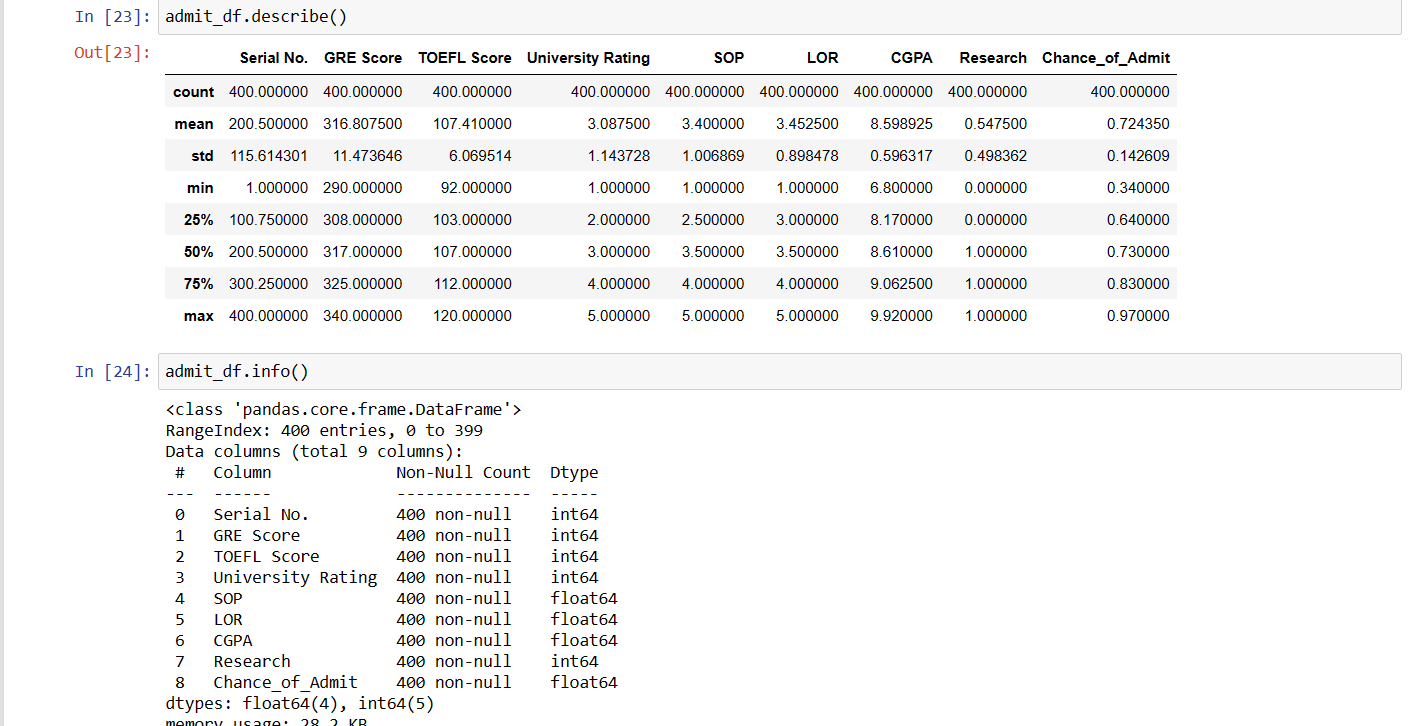
We can load the data directly from the UCI Machine Learning repository. We are using pandas to load the data. We will also use pandas next to explore the data both with descriptive statistics and data visualization.

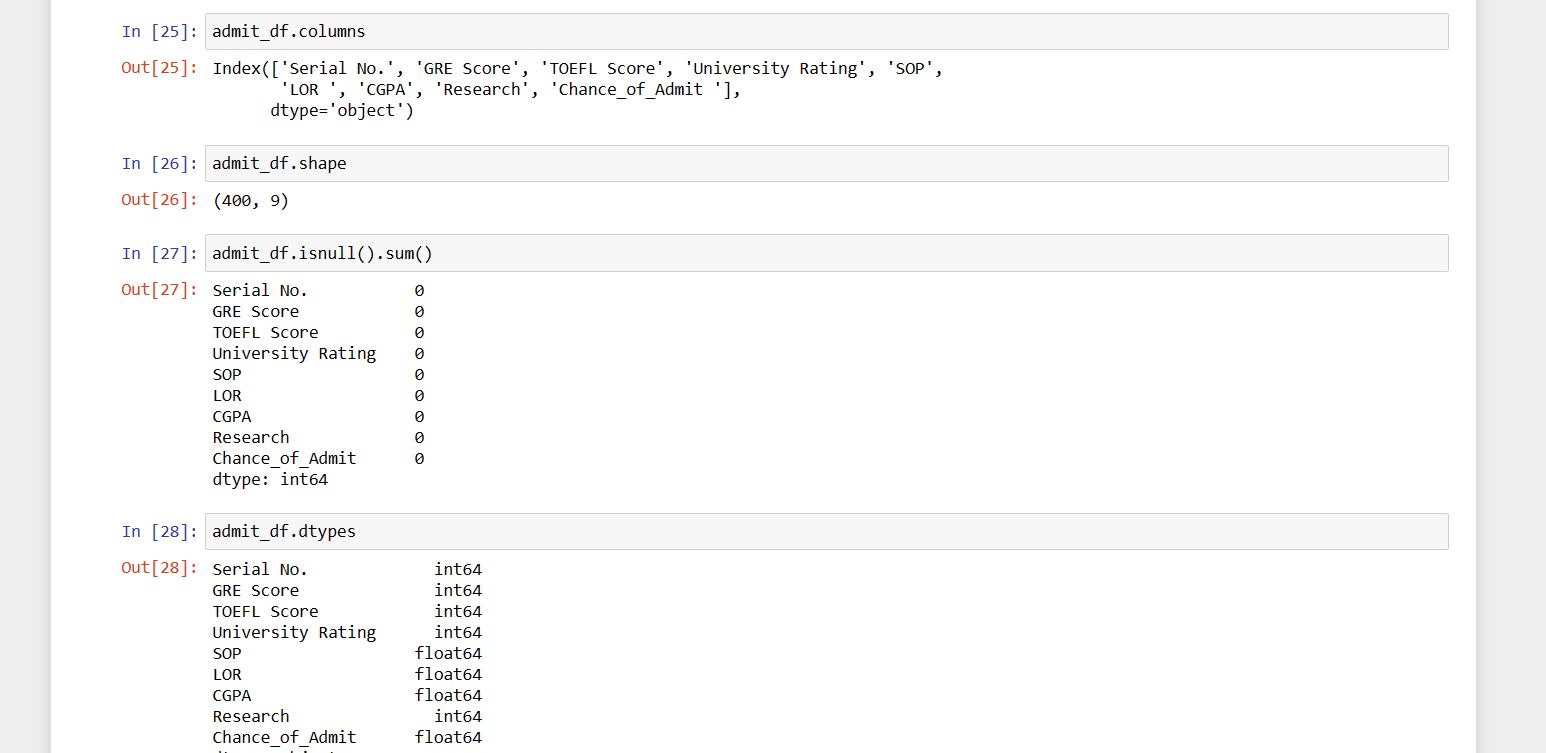


# **SUMMARIZE THE DATASET**

Now it is time to look at the data.

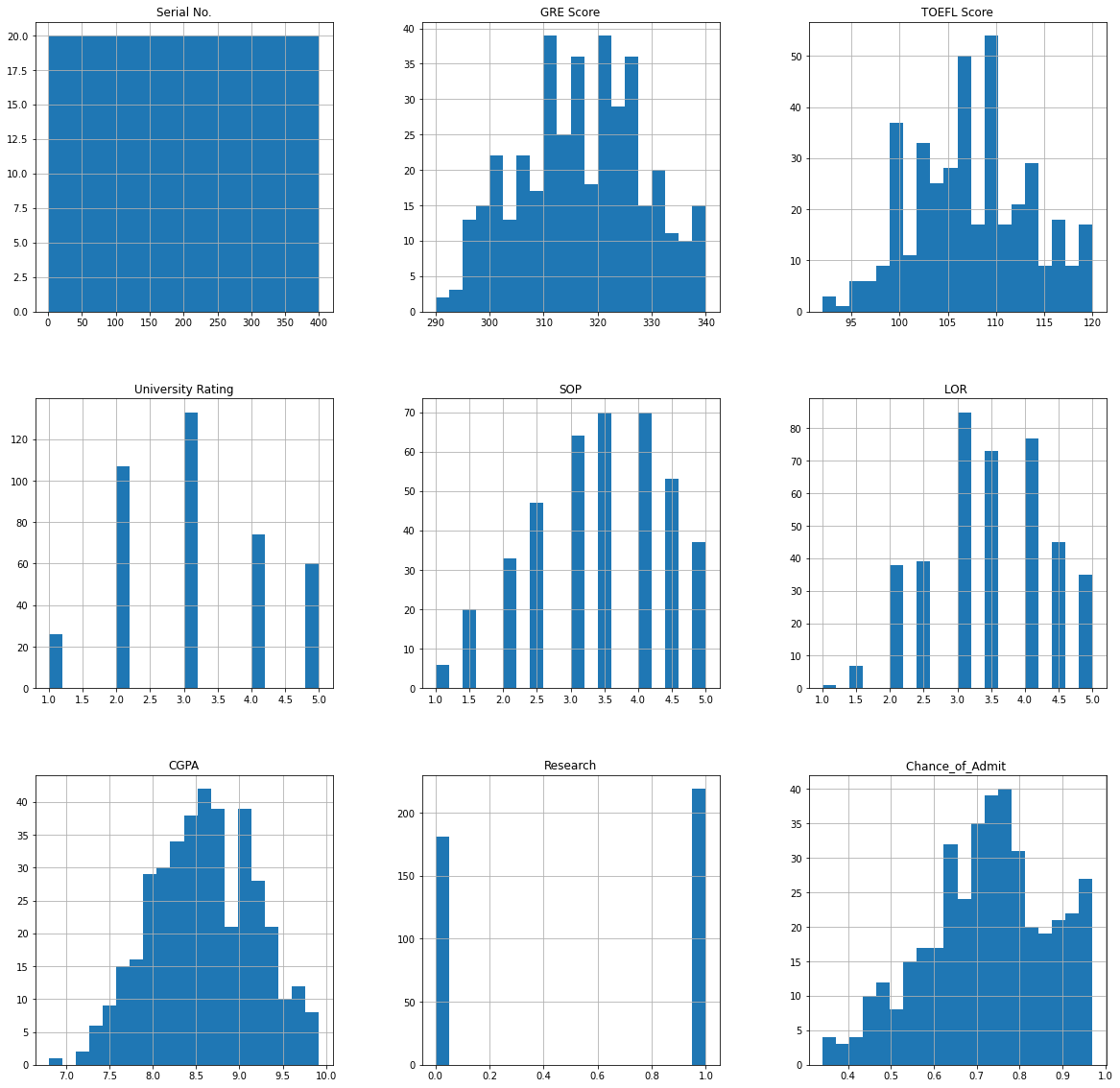
In this step we are going to look at the data a few different ways.

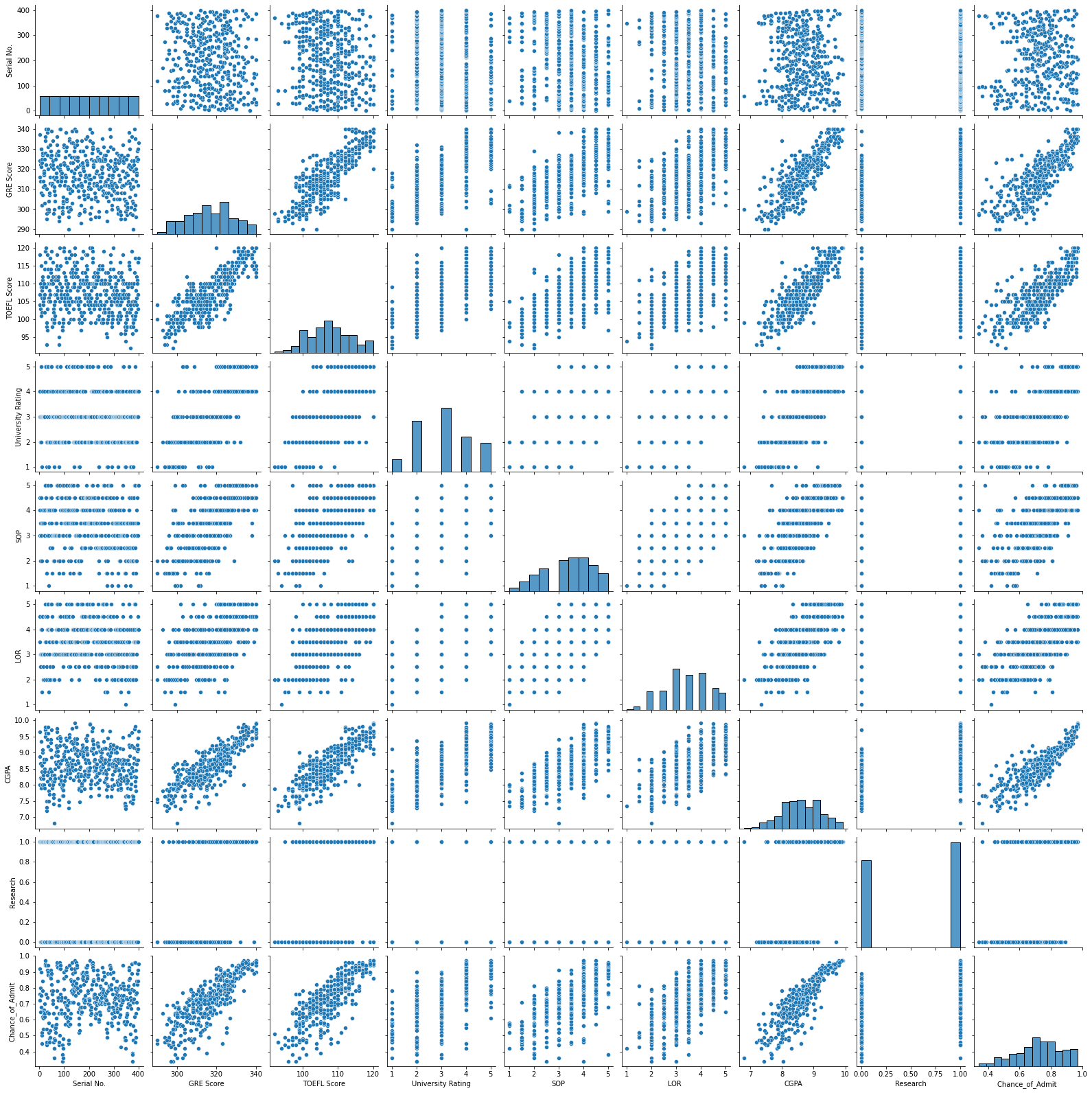


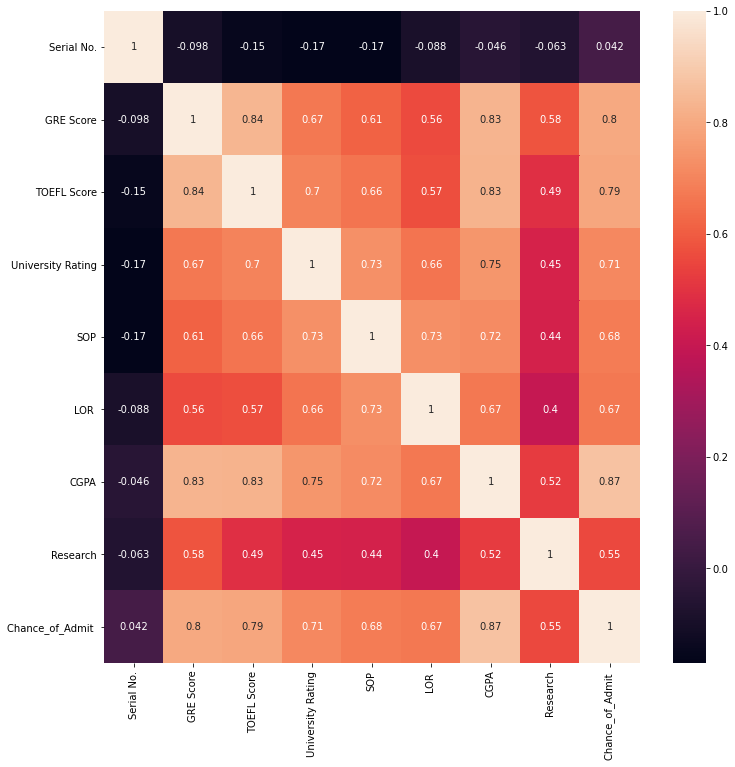


**DATA VISUALIZATION**

We now have a basic idea about the data. We need to extend that with some visualizations.







# **CONCLUSION**

Heart is the most essential organ of the human body and day by day the loss of Human Life is increasing exponentially due to heart failure. It was experimentally found that the Global pandemic Corona Virus causes heart injury among a lot of patients. Hence there is an urgent need for research to focus into the causes for heart failure and to design a robust prediction system to detect at early stage so that loss of life can be avoided.

Even though there were many heart diseases prediction systems available at present but each one has its own limitations. The main objective of this research work is to overcome the difficulty faced by other researchers and to build a robust system which works efficiently and will able to predict accurately the possibility of heart attack at very early stage.

This research work could able to design a very robust and accurate model to predict the possibility of heart failure in the current scenario. By using the Logistic Regression this model could able to predict with an accuracy of about 86.6% which is highest as compared to other algorithms.

**REFERENCES**

* <https://www.kaggle.com/andrewmvd/heart-failure-clinical-data>

* https://scikitlearn.org/stable/modules/generated/sklearn.ensemble.Ran domForestCl assifier.html

* https://stackoverflow.com/questions/38963734/

* [https://pandas.pydata.org/pandasdocs/stable/reference/api/pandas.plot ting.scatter\_matrix.html](https://pandas.pydata.org/pandas-docs/stable/reference/api/pandas.plotting.scatter_matrix.html)

* https://towardsdatascience.com/data-visualization-for-machinelearningand-data-science-a45178970be7

* https://www.datacamp.com/community/tutorials/randomforestsclassifier-python

* https://www.analyticsvidhya.com/blog/2021/10/a-comprehensiveguideon-data-visualization-in-

python/#:~:text=%20%20%201%20Bar%20Charts%0ABar%20charts%20a re,Maps%0ATemperature%20maps%20represent...%206%20Line%20Plo t%20More%20

* <https://www.geeksforgeeks.org/support-vector-machine-algorithm/>

* <https://www.educba.com/svm-algorithm/>

* [https://stackoverflow.com/questions/22533397/f1-score-in-pythoncode](https://stackoverflow.com/questions/22533397/f1-score-in-python-code)

* [https://www.springboard.com/blog/ai-machine-learning/14essentialmachine-learning-algorithms/](https://www.springboard.com/blog/ai-machine-learning/14-essential-machine-learning-algorithms/)

* [https://datascience.stackexchange.com/questions/28574/decisiontreecl assifier-object-has-no-attribute-importances](https://datascience.stackexchange.com/questions/28574/decisiontreeclassifier-object-has-no-attribute-importances)