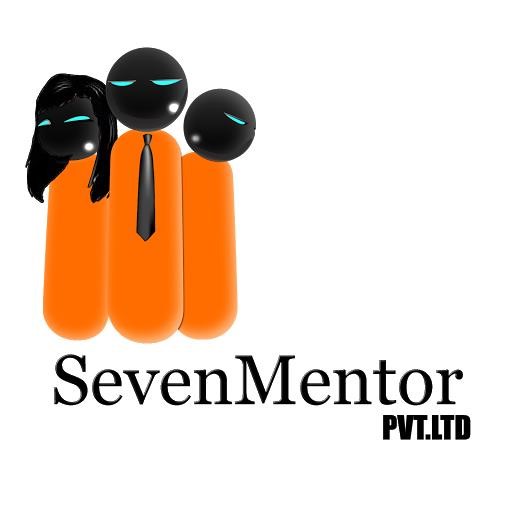
**Project File**

**Submitted by**

**Mr.**

**Pushpraj soni**



|  |  |
| --- | --- |
| **ACKNOWLEDGEMENT**    I am delighted to offer my genuine gratitude and heartfelt appreciation to everyone who has supported me in any way during my Data Analyst course. I want to use this opportunity to thank everyone. It is a sincere expression of my gratitude to them and not just a formality or ritual. The fact that I was able to learn Python, SQL and Data analysis, Seven Mentor Pvt. Ltd. in Pune makes me feel quite fortunate.  First and foremost, I want to express my gratitude to my mentor, Nishesh Sir, for his helpful guidance, unwavering support, and kindness during my course. Along with his extensive expertise and wealth of knowledge in Python, SQL and data analysis his additional attributes, such as management skills, lively attitude, focused mind set, and leadership qualities have inspired me over the course of our class and daily life.  I am grateful to all the authorities, trainers and supporting staff of Seven Mentor Pvt. Ltd. Also, I also want to thank all of my friends from every batch I've ever been in.    Thank you all.                  Mr. Nishesh Gogia Mr. Pushpraj Soni  Trainer | |
| **INDEX**         1. Introduction ………………………………………………….....4      1. Project 1- URL Convertor ………………………….…………..5   *Problem statement*  *Coding*  *Conclusion*     1. Project 2- Zomato Data Analysis………………………..……...6   *Introduction*  *Coding*  *Conclusion*     1. Project 3- email varification…………..………….... 12   *Problem statement*  *working*  *Coding*  *Conclusion*     1. Learnings ……………………………………………………...13      1. References ………………………………………………….....14 | |

**Introduction**

Machine learning (ML) is a subfield of artificial intelligence (AI) that allows computers to "learn" from training data and improve over time without being explicitly programmed.

The main aim of machine learning is to understand the structure of data and fit it into models that people can comprehend and use. Through the use of sample data or previous experience, machine learning allows computers to optimize performance standards. A model has been developed up to a certain point, and training is the process of running a computer programme to update the model's features based on training data or prior information. The model might be classified or regressive, allowing users to learn from data and make future predictions.

The Python-Machine Learning combination has cemented its place in the IT and data science industries. Python is being used by a number of market leaders for a variety of activities. According to many experts in this domain, Python's popularity is expected to skyrocket in the future years. Python's practical application in machine learning projects and activities has simplified the work of developers, data scientists, and machine learning engineers. The ecosystem of libraries, frameworks, and tools for Python is expanding rapidly. These tools and libraries include pre-written code that allows users to execute a variety of operations while saving a significant amount of time spent on code generation. Python is one of the most popular data science languages since it is simple to evaluate and combine accessible data.

Over the duration of this course, I've worked on three different projects. In order to shorten long URLs, I employed the fundamentals of object-oriented programming (OOPs) in Python. In my later studies, I worked on Zomato data analysis of Pune city using an exploratory data analysis method (EDA). My third project addressed predicting customer churn in the banking industry.

# Project 1: URL convertor

**Problem statement:**

No matter where you use it—in social media, text messages, or emails—a long URL will always look odd. But even worse, lengthy URLs have a very suspicious appearance. It has so many letters, special characters, numbers, and other things that if any one of them is misplaced or whatever, the link might do anything when we click it. Yes, having a URL that is as brief as possible is necessary to keep things appearing clean, organized, and, pleasant. It's indeed better for readability and engagement to keep things brief and to the point, even for posts like Facebook or Instagram when the character restrictions are in the thousands. Another advantage of shorter URL is that they are easier to read in chat or email, while longer links may be difficult to read or completely disrupted by line breaks.

**Working:**

I worked on this project with the use of Python's object-oriented programming (OOPs). I used the URL of KEM hospital's Google location, which appears to be somewhat lengthy.

**Coding:**



**Conclusion:** We can see that after utilizing the OOPs approach, the lengthy URL has been reduced to just 10 characters of short URL.

# Project 2: Zomato Data Analysis

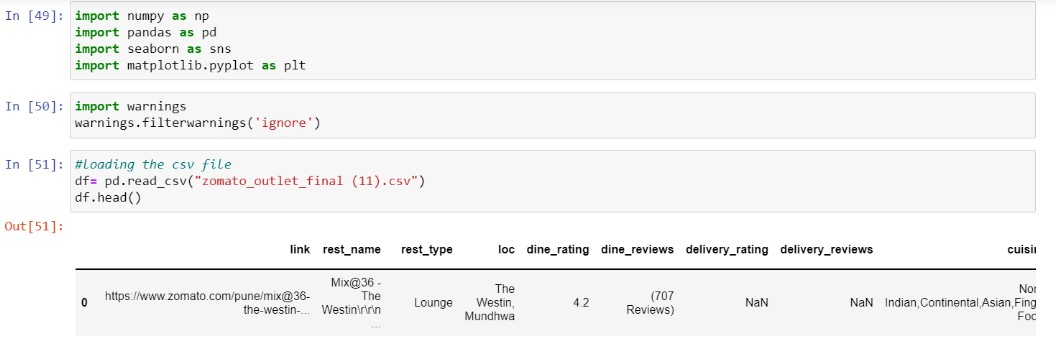
**Introduction:**

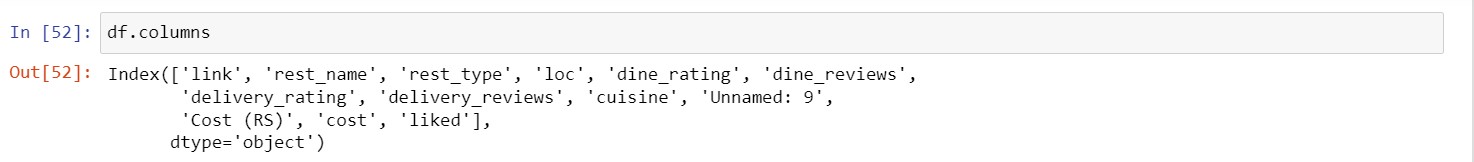
Pune is one of India's best metropolitan areas. It is also well-known for serving traditional Maharashtrian food. Regardless, if one wants to experience the other culinary culture in Pune city, it is important to explore with numerous foods, restaurants, price, location, and so on. To that purpose, I worked on this project in which I thoroughly analyzed the Zomato data, which may be able to address every question you may have on the cuisine of Pune city.

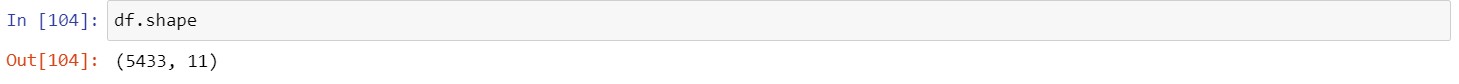
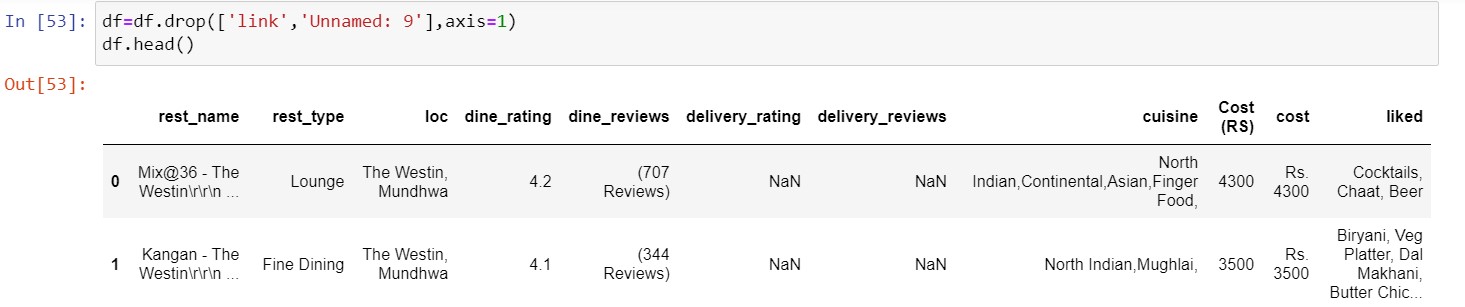
**Working:**

In order to understand the correlation between the various features in my data, I employed a number of exploratory data analysis (EDA) approaches while working on this project. I analyzed my data using both univariate and bivariate methods, and to do so, I attempted to make a number of plots, including histograms, bar plots, probability density functions (pdfs), box plots, violin plots, pie charts, scatter plots, pair plots, and so on.

**Coding:**

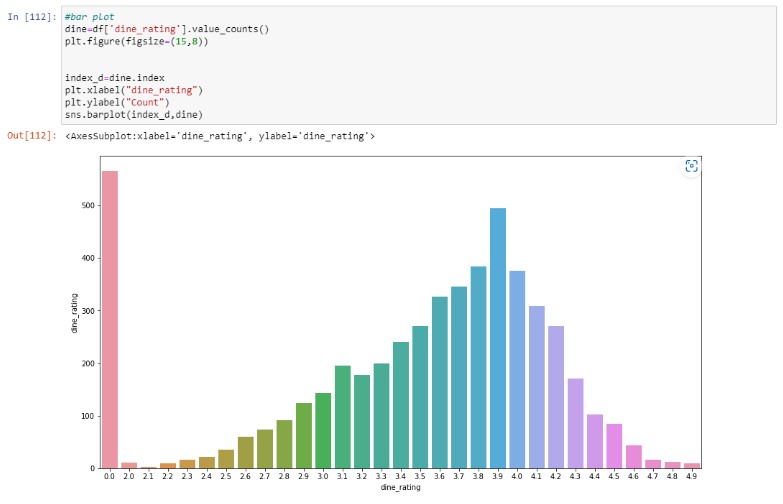
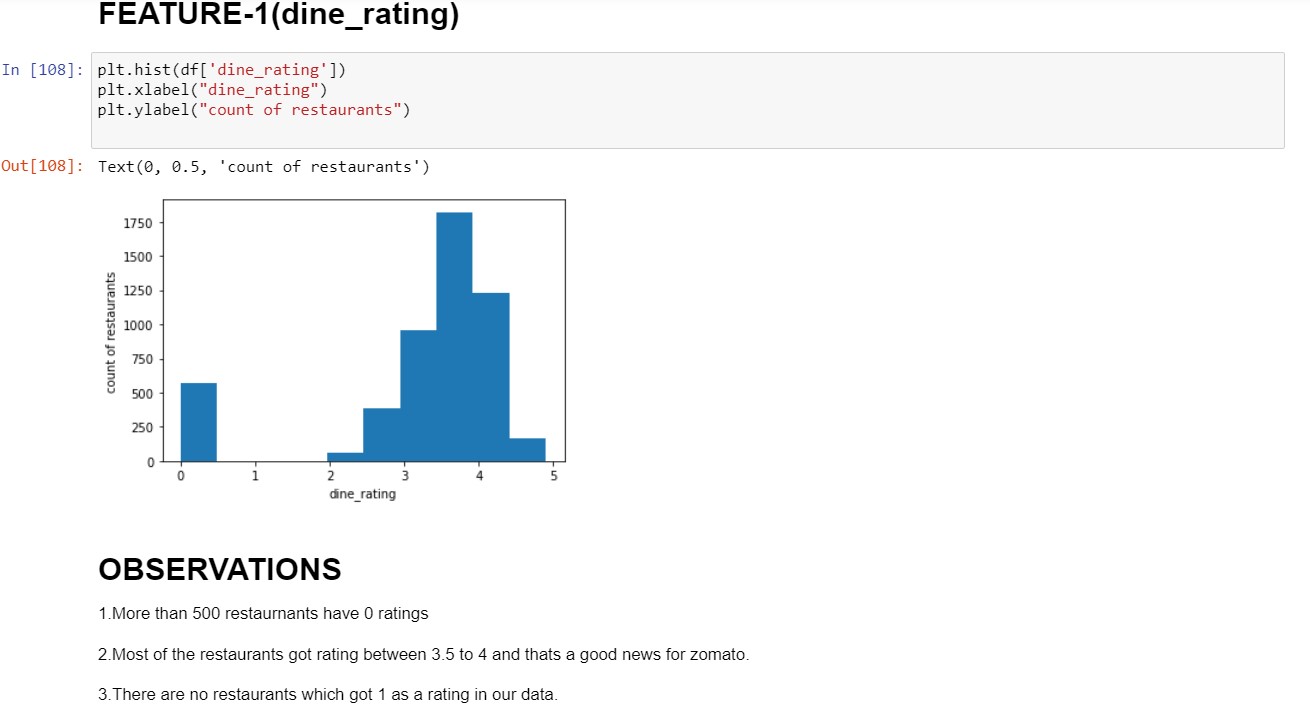






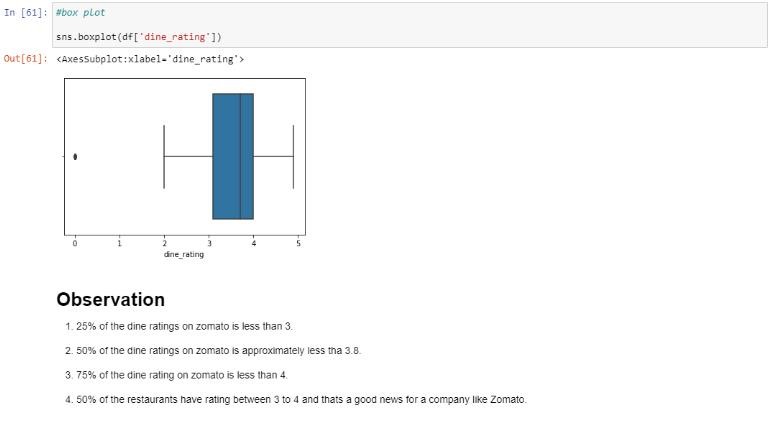
In our database, there are 5433 restaurants, including small and medium-sized eateries, bakeries, dessert bars, cake shops, bars, big restaurants, casual restaurants, etc.

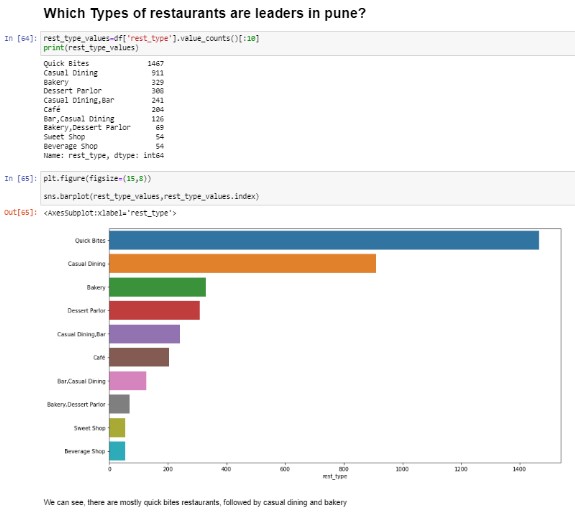
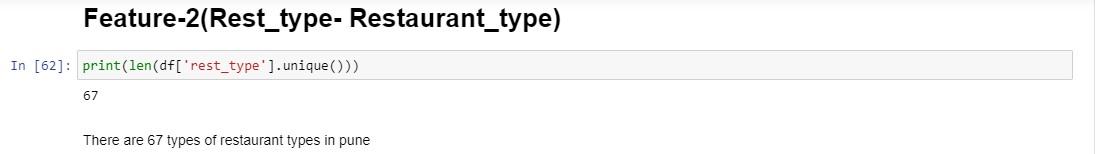
We must handle some null values in some features step by step, additionally, we have 250 duplicate rows that we must first eliminate in order to avoid errors in our results.

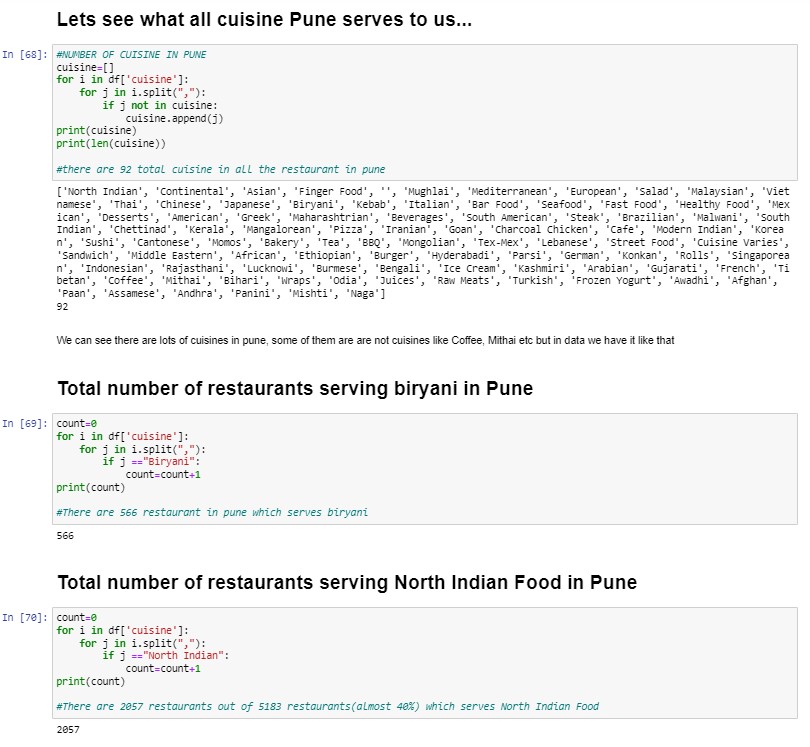


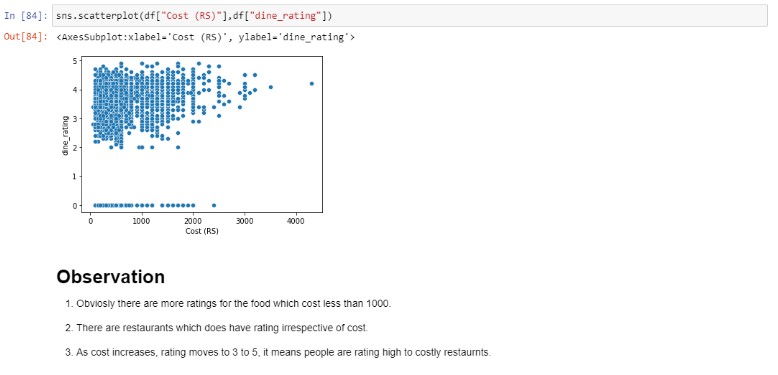


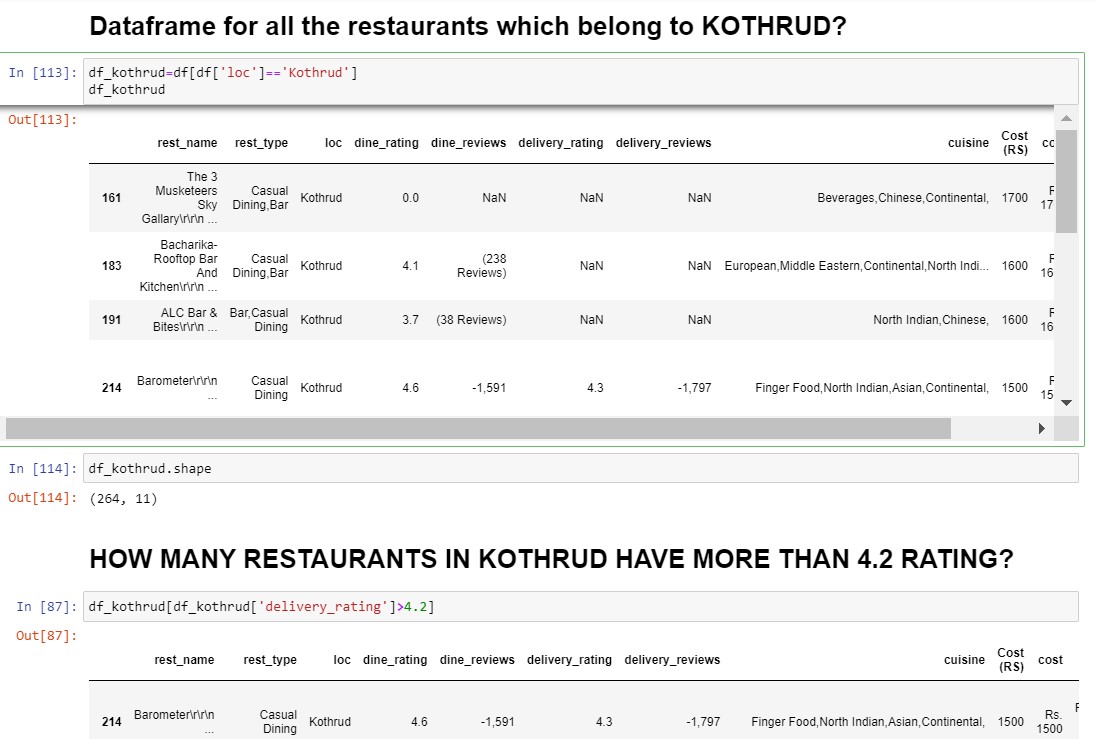
After removing duplicate rows from our dataset, we got 5183 restaurants.











**Conclusion:**

In this project, I used both univariate and bivariate exploratory data analysis approaches, analysing every feature in every way possible. Additionally, I attempted to explicate the relationship between two distinct features and how I might extract more information from it.

In addition to these results, I created a separate data frame from each prominent area in Pune and its famous food, pricing, rating, etc. In order to learn more about the food in Pune city and its specific area, I applied the EDA approach to each of these data frames and performed a thorough analysis

**Project 1: Email verifier.**

**Problem statement:**

Email is important for communication because it allows users to send information in letter format, and email can replace traditional mail options. Emails can be more beneficial for communication because they can often include text, documents and multimedia, like photos and videos. If provided email is invalid then it generated communication gap.

Valid email addresses are essential for any company hoping to succeed with email marketing and customer communication. It's critical to double-check your email address list is correct, error-free, and that the emails actually exist.

**Working:**

I worked on this project with the use of if , elif , else. I used this email verification for various email checking. which is very helpful.

**Coding:**



**Conclusion:** We can see that after utilizing the loop and conditional approach, the invalid email address has been declined on the entry time.

**Learnings:**

1. I utilised the OOPs concept in my first project to shorten long URLs, which resulted in lengthy URLs being shortened to only 10 characters.

1. In the second project, I used EDA methods to examine the relationships between various features. To do this, I used a variety of plots, including as histograms, bar plots, probability density functions (pdfs), box plots, violin plots, pie charts, scatter plots, pair plots, and so on

3. In the third project, I used if, elif ,else and loop for checking out the given email address is correct or not.

.

**References:**

http://scikit-learn.org

https://www.analyticsvidhya.com

https://pythonprogramming.net

https://courses edx.org

https://www.youtube.com

<https://www.newagepublishers.com/samplechapter/001938.pdf>

[http://mathfaculty.fullerton.edu/mathews/n2003/LogisticE](http://mathfaculty.fullerton.edu/mathews/n2003/Logistic)quationMod.html

https://www.ima.umn.edu/-tmlai/M119-Section47.pdf

<https://google.com>