**A Project Report**

**On**

***ZOMATO DATA ANALYSIS***

Submitted by:

AVINASH BORUAH

Under the guidance of

**Mr. Prabhat Das**

(PGT-IP)



**ARMY PUBLIC SCHOOL JORHAT**

**MES GATE ROWRIAH JORHAT**

**Assam-785004**

CERTIFICATE BY PRINCIPAL

This is to certify that this project report entitled “**Zomato Data Analysis**” submitted by Avinash Boruah to Army Public School Jorhat has been examined and evaluated.

The report has been prepared as per the regulations of CBSE and qualifies to be accepted.

Date:

Place:

Mrs. Firdausi Sultana Hazarika

(Principal)

Army Public School Jorhat

CERTIFICATE BY EXAMINERS

This is to certify that this project report entitled “**Zomato Data Analysis**” is the bona fide work of who carried out the project work under my supervision and guidance.

To the best of my knowledge, the matter embodied in the report has not been submitted to any other institute for the award of any other degree.

Date:

Place:

Mr. Prabhat Das

(External Examiner) (Internal Examiner)

ACKNOWLEDGEMENT

I take this opportunity to extend my heart full gratitude to Army Public School Jorhat for providing me the opportunity.

I am highly grateful to my guide Mr. Prabhat Das, PGT-IP, Army Public School Jorhat for giving us the opportunity to work under him and providing us an ample guidance and support through the project.

Lastly, I would also like to thank the authors whose publications guided us regarding our project.

DECLARATION

I admit that this report is of my own work and all the sources of the information used in this report have fully acknowledged.

I hereby declare that the dissertation work entitled “**Zomato Data Analysis**” submitted to the Army Public School Jorhat, is prepared by me and was not submitted to any other institution for award of any other degree.

Date:

Place:

Signature

Abstract

Data analysis is a process of inspecting, cleansing, transforming, and modelling data with the goal of discovering useful information, informing conclusions, and supporting decision-making.

In order to know all the features offered by the restaurants in Zomato, we’re going to analyse a sample data of the same. Features like: Number of cities present, Table option, Online delivery option, graphical representation of the ratings of the restaurants in the website, etc will be thoroughly analysed.

By analysing so, we’ll be able to grasp the huge data in a simpler manner and friendlier manner.

Contents

* Tools and Library 1
* Introduction 3
* Project overview 4
* Source code 7
* Commands in SQL 12
* Conclusion and Future work 13
* References 14 List Of Figures

Page No

* Fig 1. User information table 5
* Fig 2. Login procedure 5
* Fig 3 Output for option 1 5
* Fig 4. Output for option 2 5
* Fig 5. Output for option 3 5
* Fig 6. Output for option 4 6
* Fig 7. Output for option 5 6

Tools and Libraries Used

MySQL

MySQL is the world's most used relational [database management system](https://www.limswiki.org/index.php?title=Database_management_system&action=edit&redlink=1) (RDBMS) that runs as a server providing multi-user access to a number of databases. It is named after developer Michael Widenius' daughter. The phrase "SQL" stands for "structured query language."

[Open source projects](https://www.limswiki.org/index.php/Category:Open-source_software) that require a full-featured database management system often use MySQL. Applications which use MySQL databases include:

[TYPO3](https://www.limswiki.org/index.php/TYPO3), [Joomla](https://www.limswiki.org/index.php/Joomla), [WordPress](https://www.limswiki.org/index.php/WordPress), [phpBB](https://www.limswiki.org/index.php/PhpBB), [Drupal](https://www.limswiki.org/index.php/Drupal), and other software built on the LAMP software stack. MySQL is also used in many high-profile, large-scale Web products, including Wikipedia, Google (though not for searches), Facebook, and Twitter.

The free open-source version of MySQL is commonly referred to as the MySQL Community Edition. For commercial use, several paid editions are also available, offering additional functionality.[1]

PyCharm

PyCharm is an [integrated development environment](https://en.wikipedia.org/wiki/Integrated_development_environment) (IDE) used in [computer programming](https://en.wikipedia.org/wiki/Computer_programming), specifically for the [Python](https://en.wikipedia.org/wiki/Python_(programming_language)) language. It is developed by the [Czech](https://en.wikipedia.org/wiki/Czech_Republic) company [JetBrains](https://en.wikipedia.org/wiki/JetBrains).[[5]](https://en.wikipedia.org/wiki/PyCharm#cite_note-5) It provides code analysis, a graphical debugger, an integrated unit tester, integration with [version control systems](https://en.wikipedia.org/wiki/Revision_control) (VCSes), and supports web development with [Django](https://en.wikipedia.org/wiki/Django_(web_framework)) as well as [data science](https://en.wikipedia.org/wiki/Data_science) with [Anaconda](https://en.wikipedia.org/wiki/Anaconda_(Python_distribution)).

PyCharm is [cross-platform](https://en.wikipedia.org/wiki/Cross-platform), with [Windows](https://en.wikipedia.org/wiki/Windows), [macOS](https://en.wikipedia.org/wiki/MacOS) and [Linux](https://en.wikipedia.org/wiki/Linux) versions. The Community Edition is released under the [Apache License](https://en.wikipedia.org/wiki/Apache_License), and there is also Professional Edition with extra features – released under a [proprietary license](https://en.wikipedia.org/wiki/Proprietary_software). [2]

Pandas

In [computer programming](https://en.wikipedia.org/wiki/Computer_programming), pandas is a [software library](https://en.wikipedia.org/wiki/Software_library) written for the [Python programming language](https://en.wikipedia.org/wiki/Python_(programming_language)) for data manipulation and analysis. In particular, it offers data structures and operations for manipulating numerical tables and [time series](https://en.wikipedia.org/wiki/Time_series). It is [free software](https://en.wikipedia.org/wiki/Free_software) released under the [three-clause BSD license](https://en.wikipedia.org/wiki/3-clause_BSD_license). The name is derived from the term "[panel data](https://en.wikipedia.org/wiki/Panel_data)", an [econometrics](https://en.wikipedia.org/wiki/Econometrics) term for data sets that include observations over multiple time periods for the same individuals. Its name is a play on the phrase "Python data analysis" itself. [Wes McKinney](https://en.wikipedia.org/wiki/Wes_McKinney) started building what would become pandas at [AQR Capital](https://en.wikipedia.org/wiki/AQR_Capital) while he was a researcher there from 2007 to 2010. [3]

Numpy

NumPy (pronounced [/ˈnʌmpaɪ/](https://en.wikipedia.org/wiki/Help:IPA/English) ([NUM-py](https://en.wikipedia.org/wiki/Help:Pronunciation_respelling_key)) or sometimes [/ˈnʌmpi/](https://en.wikipedia.org/wiki/Help:IPA/English) ([NUM-pee](https://en.wikipedia.org/wiki/Help:Pronunciation_respelling_key))) is a library for the [Python programming language](https://en.wikipedia.org/wiki/Python_(programming_language)), adding support for large, multi-dimensional [arrays](https://en.wikipedia.org/wiki/Array_data_structure) and [matrices](https://en.wikipedia.org/wiki/Matrix_(math)), along with a large collection of [high-level](https://en.wikipedia.org/wiki/High-level_programming_language) [mathematical](https://en.wikipedia.org/wiki/Mathematics) [functions](https://en.wikipedia.org/wiki/Function_(mathematics)) to operate on these arrays. The ancestor of NumPy, Numeric, was originally created by [Jim Hugunin](https://en.wikipedia.org/wiki/Jim_Hugunin) with contributions from several other developers. In 2005, [Travis Oliphant](https://en.wikipedia.org/wiki/Travis_Oliphant) created NumPy by incorporating features of the competing Numarray into Numeric, with extensive modifications. NumPy is [open-source software](https://en.wikipedia.org/wiki/Open-source_software) and has many contributors. [[4]](https://en.wikipedia.org/wiki/NumPy#cite_note-Nature-5)

Matplotlib

Matplotlib is a [plotting](https://en.wikipedia.org/wiki/Plotter) [library](https://en.wikipedia.org/wiki/Library_(computer_science)) for the [Python](https://en.wikipedia.org/wiki/Python_(programming_language)) programming language and its numerical mathematics extension [NumPy](https://en.wikipedia.org/wiki/NumPy). It provides an [object-oriented](https://en.wikipedia.org/wiki/Object-oriented_programming) [API](https://en.wikipedia.org/wiki/API) for embedding plots into applications using general-purpose [GUI toolkits](https://en.wikipedia.org/wiki/GUI_toolkit) like [Tkinter](https://en.wikipedia.org/wiki/Tkinter), [wxPython](https://en.wikipedia.org/wiki/WxPython), [Qt](https://en.wikipedia.org/wiki/Qt_(software)), or [GTK+](https://en.wikipedia.org/wiki/GTK%2B). There is also a [procedural](https://en.wikipedia.org/wiki/Procedural_programming) "pylab" interface based on a [state machine](https://en.wikipedia.org/wiki/State_machine) (like [OpenGL](https://en.wikipedia.org/wiki/OpenGL)), designed to closely resemble that of [MATLAB](https://en.wikipedia.org/wiki/MATLAB), though its use is discouraged. [SciPy](https://en.wikipedia.org/wiki/SciPy) makes use of Matplotlib.

Matplotlib was originally written by [John D. Hunter](https://en.wikipedia.org/wiki/John_D._Hunter). Since then it has an active development community and is distributed under a [BSD-style license](https://en.wikipedia.org/wiki/BSD_licenses). Michael Droettboom was nominated as matplotlib's lead developer shortly before John Hunter's death in August 2012 and was further joined by Thomas Caswell.[[5]](https://en.wikipedia.org/wiki/Matplotlib#cite_note-6)

Introduction

The theme of my project is ‘ZOMATO DATA ANALYSIS’. This project is fine thought to make complex procedure of Zomato data in an easy manner which is systematic, modular designed, selective menu based user display. The modular design and constructed is very much user oriented in which user can easily understand the tools and can do edit of his own choice. The system is not any though more and does not possesses many application but it is made by focusing on maintaining record employee’s action in a computerized rather than time taking and cumbersome manual system.

The project software application that can easily handle by minimum educated and simple computer knowledge person without any option of error

Project Overview

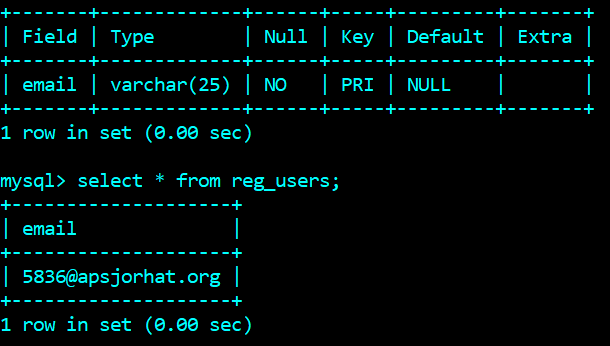


Figure 1**:**All the information regarding the users are stored in this table

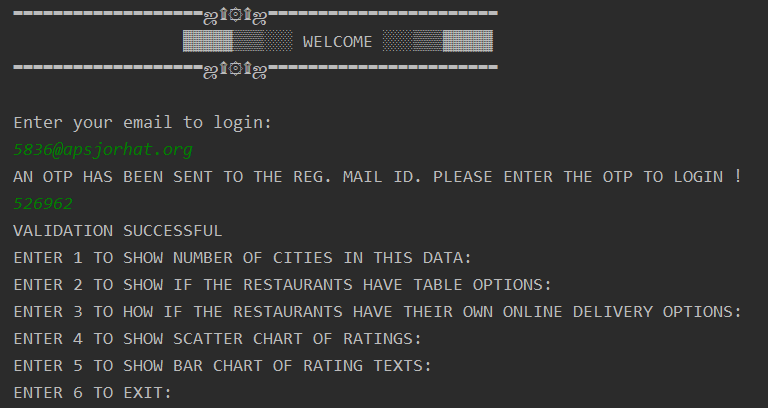
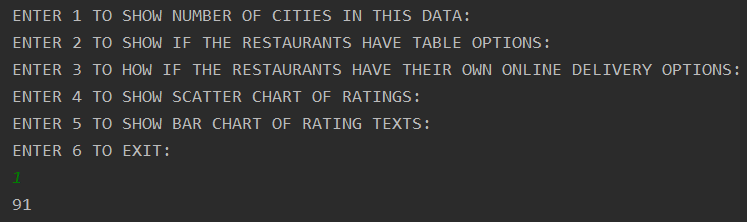


Figure 2: Login procedure for users

  
Figure 3: On entering option 1 we get the following output

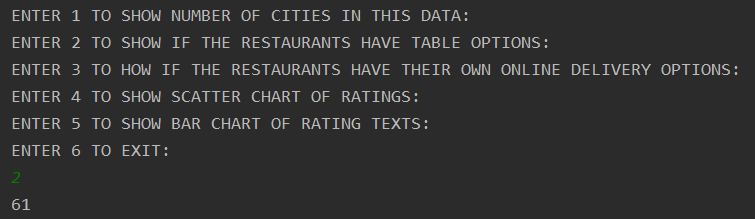


Figure 4: On entering option 2 we get the following output

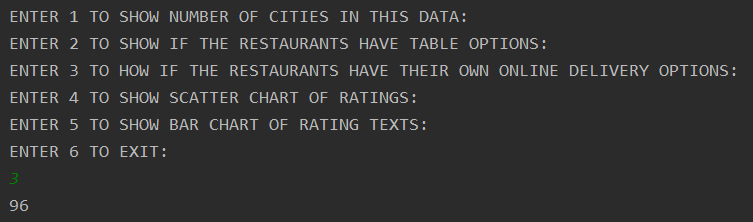


Figure 5: entering option 3 we get the following output

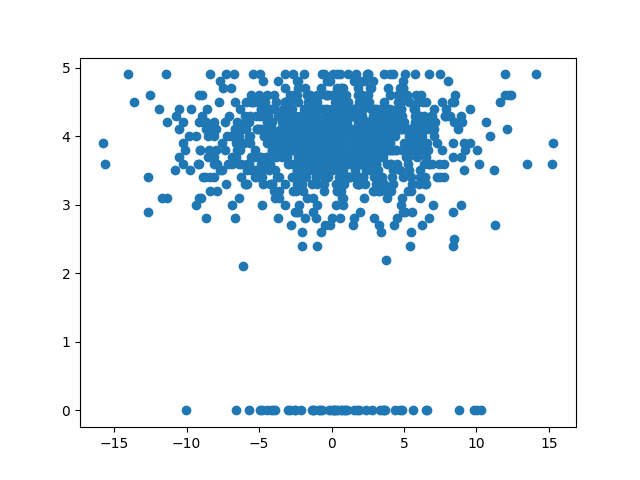


Figure 6: entering option 4 we get the following output

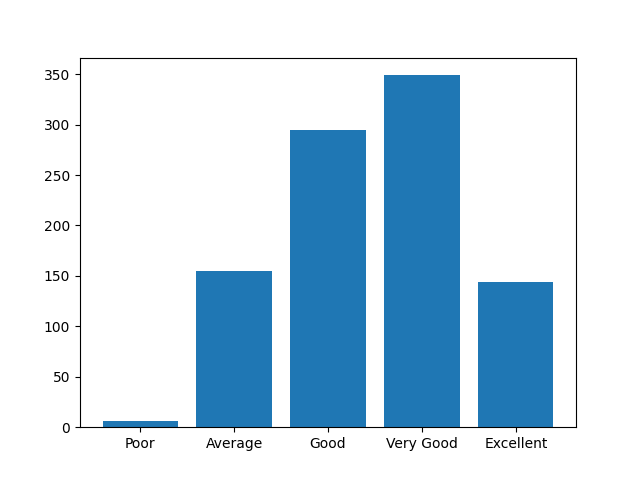


Figure 7: entering option 5 we get the following output

**Source Code**

**Auth**

import connector as con  
def auth\_user(email):  
 count=0  
 query = "select \* from reg\_users;"  
 con.cursor.execute(query)  
 for i in con.cursor:  
 if i[0]==email:  
 count=1  
 return count

**Connector**

import mysql.connector as mc  
dbc = mc.connect(host="localhost",user="root",passwd="root",database="zomato\_data")  
cursor=dbc.cursor()

**Main**

import auth  
import otp\_sender  
import dashboard  
print("▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬ஜ۩۞۩ஜ▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬\n ▓▓▓▓▓▒▒▒░░░ WELCOME ░░░▒▒▒▓▓▓▓▓\n▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬ஜ۩۞۩ஜ▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬▬\n\nEnter your email to login: ")  
email=input()  
if auth.auth\_user(email)==1:  
 rcv\_otp=otp\_sender.otp\_sender(email)  
 print("AN OTP HAS BEEN SENT TO THE REG. MAIL ID. PLEASE ENTER THE OTP TO LOGIN !")  
 inp\_otp=input()  
 if rcv\_otp == inp\_otp:  
 print("VALIDATION SUCCESSFUL")  
 dashboard.options()  
  
 else:  
 print("INVALID OTP")  
 exit()  
else:  
 print("USER NOT REGISTERED!")  
 print("Please send an email to group5@apsjorhat.org to register")  
 exit()

**Dashboard**

import textblob as tb

import pandas as pd

import matplotlib.pyplot as plt

import numpy as np

import csv

# How many cities are there?

# How many restaurants have Table option

# How many has the option of online delivery via their own app

# Scatter chart of Aggregate rating

# Bar chart of Text Rating

# Exit

df= pd.read\_csv('zomato.csv', encoding='iso-8859-1', nrows=1000, on\_bad\_lines = 'warn')

def cities():

groups=df.groupby(by='City')

print(len(groups.groups.keys()))

def has\_table():

val=df[df['Has Table booking']=='Yes']

print(val['City'].count())

def delivery():

val\_online=df[df['Has Online delivery']=='Yes']

print(val\_online['City'].count())

def ratings():

ratings=list(df['Aggregate rating'])

x = np.random.normal(min(ratings), max(ratings), len(ratings))

plt.scatter(x, ratings)

plt.savefig("scatter\_zomato.pdf")

plt.show()

def rating():

count\_ratings=list(df['Rating text'])

Excellent=(count\_ratings.count('Excellent'))

Very\_good=(count\_ratings.count('Very Good'))

Average=(count\_ratings.count('Average'))

Good=(count\_ratings.count('Good'))

Poor=(count\_ratings.count('Poor'))

x=['Poor','Average', 'Good', 'Very Good', 'Excellent']

y=[Poor, Average, Good, Very\_good, Excellent]

plt.bar(x,y)

plt.savefig("bar\_zomato\_review.pdf")

plt.show()

def exit():

exit()

def options():

print("ENTER 1 TO SHOW NUMBER OF CITIES IN THIS DATA:\r\nENTER 2 TO SHOW IF THE RESTAURANTS HAVE TABLE OPTIONS:\r\nENTER 3 TO HOW IF THE RESTAURANTS HAVE THEIR OWN ONLINE DELIVERY OPTIONS:\r\nENTER 4 TO SHOW SCATTER CHART OF RATINGS:\r\nENTER 5 TO SHOW BAR CHART OF RATING TEXTS:\r\nENTER 6 TO EXIT:")

inp=int(input())

if inp==1:

cities()

options()

if inp==2:

has\_table()

options()

if inp==3:

delivery()

options()

if inp==4:

ratings()

options()

if inp==5:

rating()

options()

if inp==6:

pass

**otp\_sender**

import smtplib

import random

def otp\_sender(email):

otp=str(random.randint(100000,999999))

SUBJECT = 'OTP FOR LOGIN'

TEXT = 'YOUR OTP TO LOGIN IS:' + otp

s = smtplib.SMTP('smtp.gmail.com', 587)

s.starttls()

s.login('5836@apsjorhat.org', '12345678')

message = 'Subject:{} \n\n{}'.format(SUBJECT, TEXT)

s.sendmail('5836@apsjorhat.org', email, message)

s.quit()

return otp

**Commands Used In MySql**

**Creating database**

Create database zomato\_data**;**

**Using database**

Use zomato\_data**;**

**Creating table and inserting values**

Create table reg\_user(email\_id varchar(30) primary key)**;**

**Desc table**

Desc reg\_user**;**

**Inserting values**

insert into reg\_users values(“5836@apsjorhat.org”)**;**

**To fetch all values**

Select \* from reg\_user**;**

**CONCLUSION AND FUTURE WORK**

By making this project we have analysed the data of Zomato , by doing so we have been able to device a method which can help us to analyse the overall features offered by the restaurants registered in Zomato.

The functions performed by this experimental project are accurate only in accordance to our assumptions for further up gradation we can also add features such as real time analysis and a Graphical User Interface to this project, which can be developed in the form of a web , desktop or mobile application. This project can also be scaled to analyse global data in real time.

**References**

[1] <https://www.limswiki.org/index.php/MySQL>

[2] <https://en.wikipedia.org/wiki/PyCharm>

[3] <https://en.wikipedia.org/wiki/Pandas_(software)>

[4] <https://en.wikipedia.org/wiki/NumPy>

[5] <https://en.wikipedia.org/wiki/Matplotlib>