

PySpark

Mini Project

Swiggy Restaurant Rating Analysis.



1 Table of Contents

1	Introduction	2
1.1	Setup Checklist for Mini Project	3
1.2	Instructions	3
2	Problem Statement	4
2.1	Objective	4
2.2	Abstract of the project	4
2.3	Technology used:	4
3	Implementation	5
3.1	Summary of the functionality to be built:	5
3.2	Guidelines on the functionality to be built:	6
3.3	3.3 Data Model:	11
4	Reports to be built	12
5	Appendix (source Files)	15

1 Introduction

This document outlines a mini project. The project is to develop **Swiggy Restaurants Reporting System**.

This document contains the work flow of the system and gives guidelines on how to build the functionality gradually.

1.1 SETUP CHECKLIST FOR MINI PROJECT

- Hardware:
 - Intel Pentium 90 or higher (P166 recommended)
 - Microsoft Windows 2010 or above.
 - Memory: 4GB of RAM (4GB or more recommended)
- Software:
 - Anaconda
 - Python
 - IDE – Pycharm/Jupyter Notebook
 - Pyspark
 - Internet Explorer 10.0 or higher

NOTE: Anaconda will install Python. Other tools like Jupyter notebook, Spyder can be installed through Anaconda as well.

1.2 INSTRUCTIONS

- The code modules in the mini project should follow all the coding standards.
- Create a directory by your name in drive **<drive>**. In this directory, create a subdirectory **MiniProject**. Store your Project here.
- You can refer to your course material.
- You may also look up the help provided in the PYSPARK docs and documentation provided with respective tools.

2 Problem Statement

2.1 OBJECTIVE

Development of a **Swiggy Restaurant Rating Analysis**.

2.2 ABSTRACT OF THE PROJECT

The culinary in India is as varied as its culture. On every 10 kilometers of distance we find the variety in food. With increasing technology we order food online. Amongst all the food applications Swiggy is one which is very widely used.

The purpose of this project is to analyze which restaurant is popular in a particular locality of a particular city. It collects the data on various attributes like City, Cuisines, rating in number, rating in text, Average cost of two person, availability of table and online booking. There are 13 prime cities of India, 226 restaurants with 28 different cuisines spread across 137 different localities of these cities.

2.3 TECHNOLOGY USED:

- Anaconda
- Python
- IDE – Pycharm/Jupyter Notebook
- Pyspark

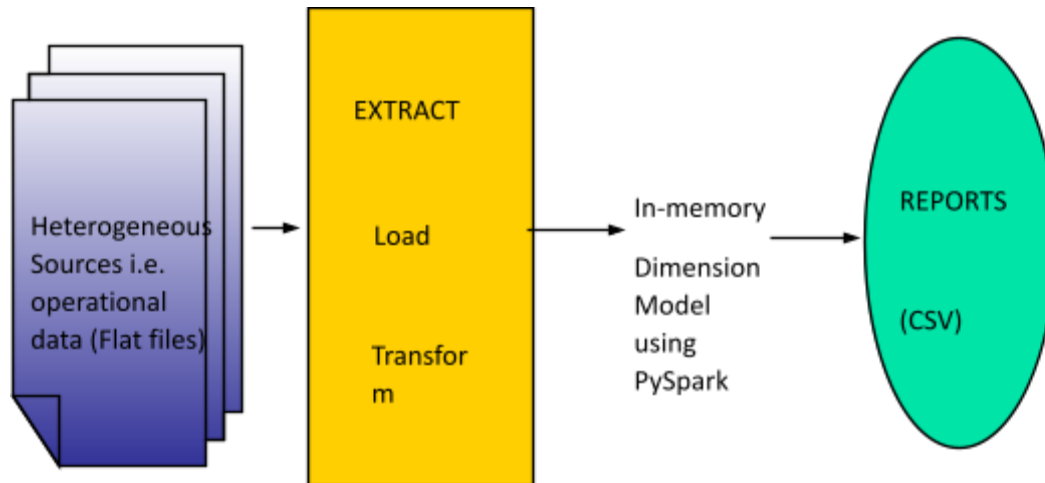
3 Implementation

3.1 SUMMARY OF THE FUNCTIONALITY TO BE BUILT:

The participants need to develop the Swiggy Rating Analysis by building the functionality incrementally in each of the course modules of PYSPARK LOT using one of the ETL and reporting tool.

3.2 GUIDELINES ON THE FUNCTIONALITY TO BE BUILT:

Project flow



FINAL_FACT	
Fact_ID	Number(p,s)
City_ID	Number(p,s)
Locality_ID	Number(p,s)
Rest_ID	Number(p,s)
Cuisine_ID	Number(p,s)
Average_cost_for_two	Number(p,s)
Rating_ID	Number(p,s)
Delivery_ID	Number(p,s)
Table_booking_ID	Number(p,s)
Votes	Number(p,s)
Price_range	Number(p,s)

Schema Design:

The project follows the start schema approach.

■ **The Dimension tables are :**

- City
- Locality
- Restaurant
- Cuisines
- Delivery
- Table booking
- Rating

■ **Fact Table is :**

- Fact_Swiggy

LOCALITY_DIM	
Locality_ID	Number(p,s)
Locality_name	Number(p,s)
CITY_DIM	
City_ID	Number(p,s)
City_name	Number(p,s)

RESTAURANT_NAME_DIM	
Rest_ID	Number(p,s)
Rest_name	Number(p,s)

CUISINE_DIM	
Cuisines_ID	Number(p,s)
Cuisine_name	Number(p,s)

CUISINE_DIM	
Cuisines_ID	Number(p,s)
Cuisine_name	Number(p,s)
TABLE_BOOKING_DIM	
Table_booking_ID	Number(p,s)
Table_booking_ava il	Varchar

RATING_DIM	
Rating_id	Number(p,s)
Rating_in_star	Number(p,s)
Ratin_in_text	Varchar

■ **City:**

This dimension contains information about the City on which the analysis will be carried out.

This CSV file comprises of the following Fields:

- ❖ City_ID
- ❖ City_Name

There are 13 Cities namely Bangalore, Chandigarh, Chennai, Goa, Guwahati, Hyderabad, Jaipur, Kolkata, Lucknow, Mumbai, Patna, Pune and Surat.

■ **Locality:**

This dimension contains information about the locality on which the analysis will be carried out.

This CSV file comprises of the following Fields:

- ❖ Locality_ID
- ❖ Locality_Name

There are 137 localities in 13 cities.

■ **Restaurant**

This dimension contains Restaurants with its name and ID:

- ❖ Restaurant_ID
- ❖ Restaurant_Name

There are 226 restaurants in 13 cities with various sub-branches and cuisines.

■ **Cuisine**

This dimension contains Cuisines available in various restaurants with its name and ID :

- ❖ Cuisine_ID
- ❖ Cuisine_Name

There are 28 different cuisines in the dataset to carry out the analysis. The Cuisines are:

- ✓ American
- ✓ Asian
- ✓ Bakery
- ✓ Bengali
- ✓ Biryanis
- ✓ Café
- ✓ Continental
- ✓ Desserts
- ✓ European
- ✓ Fast Food
- ✓ Healthy Food
- ✓ Hyderabadi
- ✓ Italian
- ✓ Japanese
- ✓ Mediterranean
- ✓ Mexican
- ✓ Modern-Indian
- ✓ North Indian
- ✓ Rajasthani
- ✓ Sea Food
- ✓ South Indian
- ✓ Street Food

■ **Table_booking:**

This dimension contains information about the Table booking availability on which the analysis will be carried out.

- ❖ Table_booking_id
- ❖ Table_booking_avail

■ **Delivery:**

This dimension contains information about the Delivery on which the analysis will be carried out.

- ❖ Delivery_ID
- ❖ Delivery_avail

■ **Rating:**

This dimension contains information about the Rating on which the analysis will be carried out.

- ❖ Rating_ID
- ❖ Rating_in_stars
- ❖ Rating_in_text

The Fact Table is:

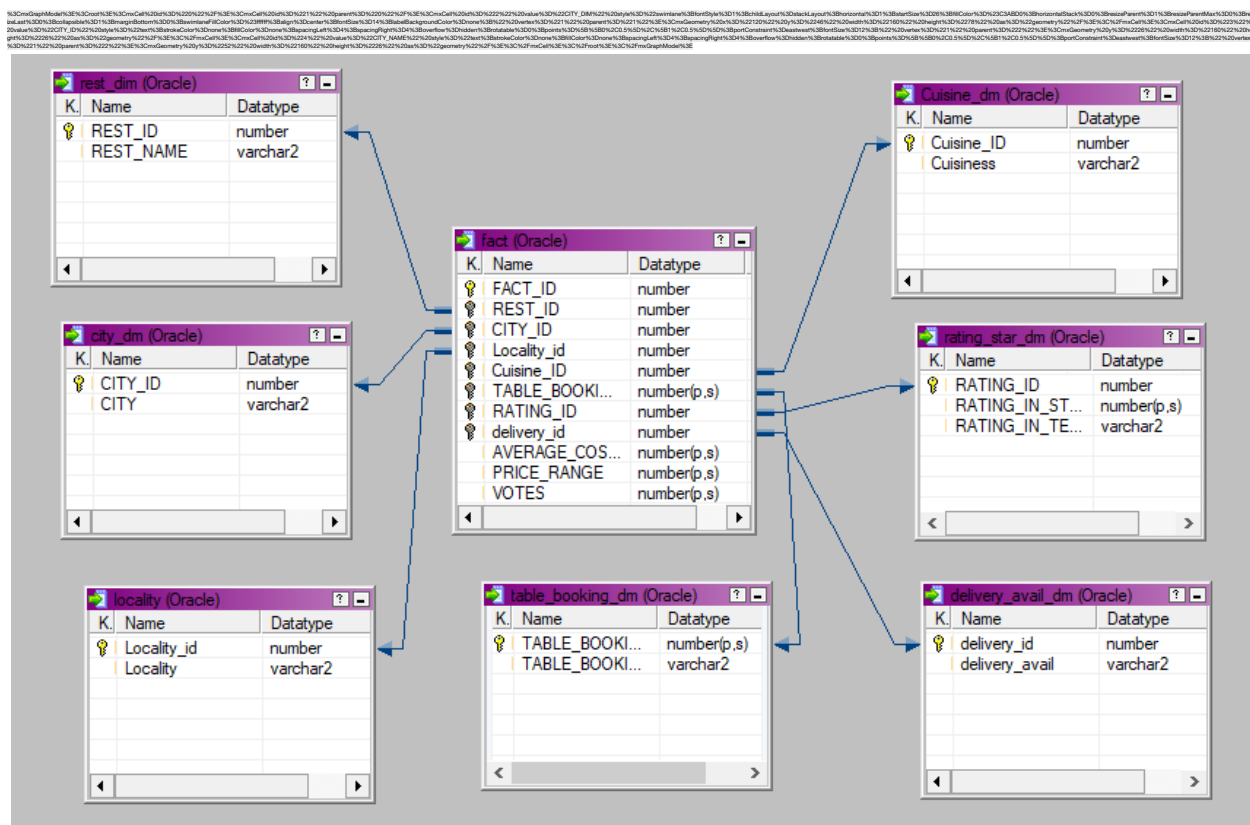
■ **Fact_Swiggy**

This is a Fact table that contains the IDs of all the cities, restaurants, locality and cuisine along with the cost for two people, ratings and availability of tables and online delivery

- ✓ Fact_ID
- ✓ City_ID
- ✓ Locality_ID
- ✓ Rest_ID
- ✓ Cuisine_ID
- ✓ Average_cost_for_two
- ✓ Rating_ID
- ✓ Delivery_ID
- ✓ Table_booking_ID
- ✓ Votes
- ✓ Price_range

3.3 DATA MODEL:

This is for reference purpose.



Note: The datatype/length for the Dimension/fact table attributes can be changed as required. Additional fields can be added, if required.

Data Transformation for data warehouse:

- 1) Load the Data into Dimension tables using the CSV files provided.
- 2) Load the Data into Fact tables using the CSV files provided.
- 3) Populate Dimension tables before fact table
- 4) Dimension table should have unique values with unique system generated IDs. Rank() function can be used to generate the unique ID (sequence). Same for Fact ID
- 5) Other ID columns in Fact table example in fact_Swiggy table – rest_id, city_id etc. column should be populated with id columns values from corresponding dimension table

4 Reports to be built

Following reports to be created. Create CSV file output for each report. Reports to be generated from the Fact and Dimension tables

Count of Restaurant city wise

City	count
Pune	20
Chandigarh	18
Lucknow	21
Mumbai	20
Patna	20

City-wise Top 10 restaurant based on Avg cost for 2

City	Restaurant Name	Average Cost for two
Mumbai	145 Kala Ghoda	1500
Mumbai	Farzi Cafe	1500
Mumbai	SpiceKlub	1500
Mumbai	The English Department Bar & Diner	1500
Mumbai	Mirchi And Mime	1500
Mumbai	R' ADDA	1200
Mumbai	Grandmama's Cafe	1100
Mumbai	Cafe Hydro	1000
Mumbai	The Fusion Kitchen	1000
Mumbai	Tea Villa Cafe	1000

Top 10 restaurant based on Avg Votes

City	Restaurant Name	Votes
Pune	18 Degrees Resto Lounge	474
Pune	Kargo	244
Pune	Le Plaisir	214
Pune	Chili's	208
Pune	Barbeque Ville	201
Pune	German Bakery Wunderbar	166
Pune	The Sassy Spoon	144
Pune	Teddy Boy	142
Pune	Barbeque Nation	136
Pune	Effingut Brewerkz	128

Top 10 restaurant based on Rating City wise

City	Restaurant Name	Rating
Bangalore	Toit	4.8
Bangalore	ECHOES Koramangala	4.7
Bangalore	Truffles	4.7
Bangalore	The Fatty Bao - Asian Gastro Bar	4.7
Bangalore	AB's - Absolute Barbecues	4.6
Bangalore	Onesta	4.6
Bangalore	Big Brewsky	4.5
Bangalore	Flechazo	4.4
Bangalore	Eat Street	4.3
Bangalore	Communiti	4.2

Rating based on delivery availability City wise

City	Online delivery	Rating
Bangalore	Yes	4.1
Bangalore	No	3.9
Hyderabad	Yes	3.9
Hyderabad	No	3.8
Chennai	Yes	3.8
Chennai	No	4

Avg cost for 2 based on cuisine city wise

City	Cuisine	Average Cost for two
Bangalore	Modern Indian	1500
Bangalore	Bakery	800
Bangalore	Fast Food	600
Bangalore	North Indian	400
Bangalore	Continental	1200