

Reducing Food Wastage by Analyzing Restaurants Takeaway Orders Dataset

Locations of important files and datasets:

- Datasets can be found under the **Datasets** folder
- Jupyter Notebook runnable file is inside **ipynb-Jupyter Notebook** folder
- Power BI runnable file can be found inside **pbix-Power BI** folder.

How to run the ipynb file:

- Install Anaconda Navigator from here:
<https://sparkbyexamples.com/python/how-to-install-anaconda-on-windows>
Installation guide for Anaconda Navigator: <https://docs.anaconda.com/anaconda/install/>
- Install findspark, pyspark, and Java from this link:
<https://sparkbyexamples.com/spark/apache-spark-installation-on-windows/>
installation tutorial for spark: <https://sparkbyexamples.com/spark/apache-spark-installation-on-windows/>
- From Anaconda Navigator open Jupyter Notebook. After that open the directory where the ipynb file is saved. Now load the file in the Jupyter Notebook text editor.
- Download MongoDB from here: <https://www.mongodb.com/docs/manual/installation/>
- Download Studio 3T from here:
<https://studio3t.com/download>
- Connect Studio 3T with MongoDB using this guide: <https://studio3t.com/knowledge-base/articles/connect-to-mongodb/>
- Download PowerBI from this link: <https://powerbi.microsoft.com/en-au/getting-started-with-power-bi/>
- In section 3 of ipynb file, put the directory where restaurant-1-orders and restaurant-2-orders are saved.

```
In [3]: # Load csv dataset
db_load = spark.read.csv(r"E:\CS4010-Data Management\Project\restaurant-1-orders.csv", header=True)
db_load2 = spark.read.csv(r"E:\CS4010-Data Management\Project\restaurant-2-orders.csv", header=True)
db_load.write.format('mongo').mode('overwrite')\
.option('spark.mongodb.output.uri', 'mongodb://127.0.0.1:27017/food_wastage.restaurant-1-orders').save()
db_load2.write.format('mongo').mode('overwrite')\
.option('spark.mongodb.output.uri', 'mongodb://127.0.0.1:27017/food_wastage.restaurant-2-orders').save()
```

- In section 12, put the directory where restaurant-1-ingredients and restaurant-2-ingredients datasets are saved.
- In section 19 and 20, put your desired local directory where you want to save the restaurant-1-details and restaurant-2-details datasets.

In [17]: # Loading Ingredients into main DF restaurant-1

```
for i in range(len(n_db)):
    for j in range(len(db_new)):
        if db_new['Item Name'][j] == n_db['Item Name'][i]:
            db_new['Ingredients'][j] = n_db['Ingredients'][i]
```

In [18]: # Loading Ingredients into main DF restaurant-2

```
for i in range(len(n_db2)):
    for j in range(len(db_new2)):
        if db_new2['Item Name'][j] == n_db2['Item Name'][i]:
            db_new2['Ingredients'][j] = n_db2['Ingredients'][i]
```

In [19]: # Saving New DF restaurant-1 (Locally)

```
restaurant_1_details = pd.DataFrame(db_new)
restaurant_1_details.to_csv("E:\\CS4010-Data Management\\Project\\restaurant-1-details.csv", header = True, index = False)
```

In [20]: # Saving New DF restaurant-2 (Locally)

```
restaurant_2_details = pd.DataFrame(db_new2)
restaurant_2_details.to_csv("E:\\CS4010-Data Management\\Project\\restaurant-2-details.csv", header = True, index = False)
```

- In section 21, put the directory where you saved restaurant-1-details and restaurant-2-details datasets.

In [21]: # # Loading New DF with ingredients

```
df = pd.read_csv("E:\\CS4010-Data Management\\Project\\restaurant-1-details.csv")
display(df)

df2 = pd.read_csv("E:\\CS4010-Data Management\\Project\\restaurant-2-details.csv")
display(df2)
```

- **Tips:** You can skip by commenting out section 17-20 as it will take a lot of time for data entry. You can use the attached dataset named **restaurant-1-details** and **restaurant-2-details** for loading the restaurant orders with ingredients. However, you must change the directory according to your saved location.

Load pbix file:

- To load the Power BI file, go to **pbix-Power BI** folder and open the pbix file.

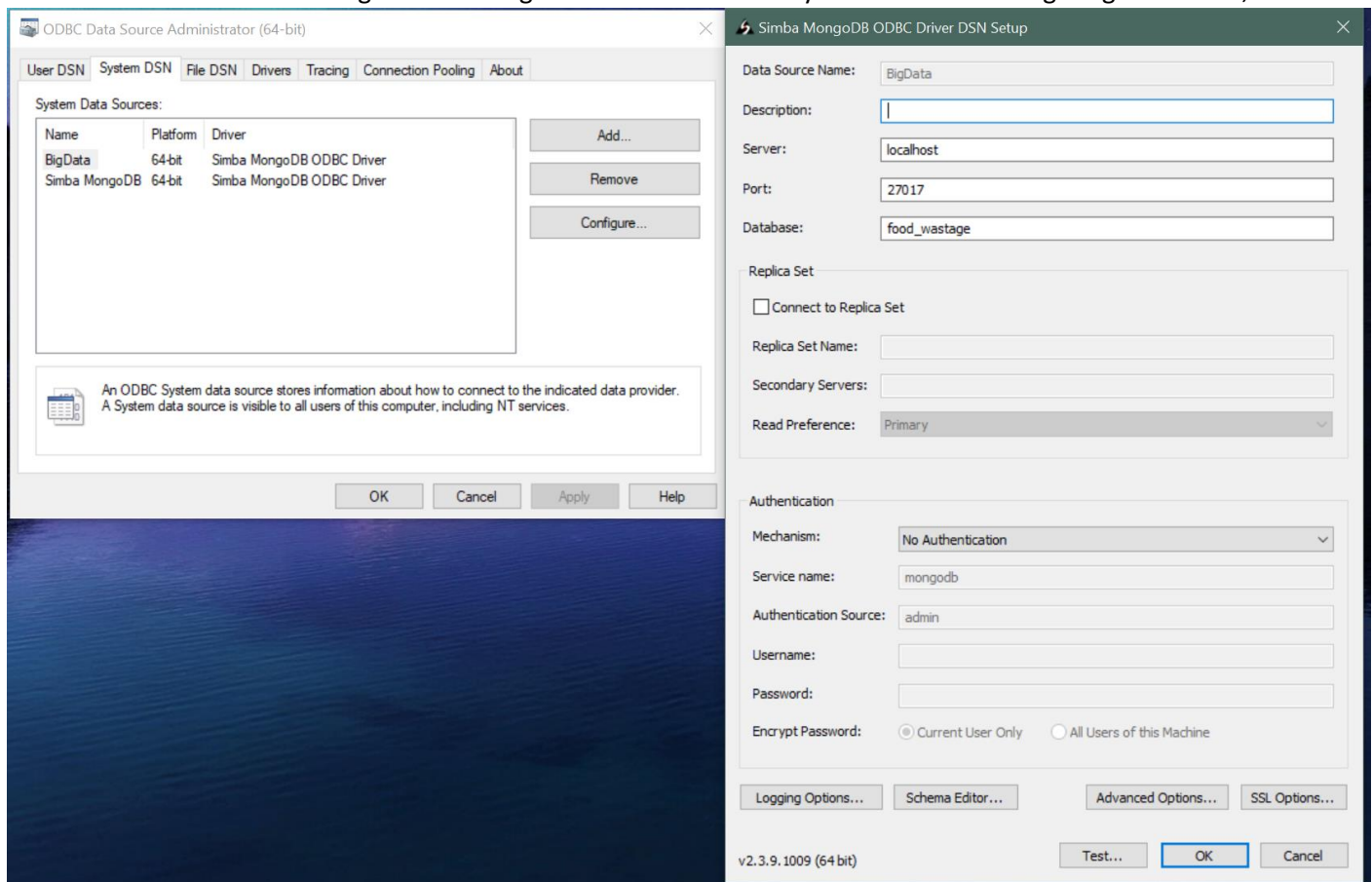
Load pbix file from scratch:

This is the procedure of how we connect our Power BI with MongoDB (Studio 3T).

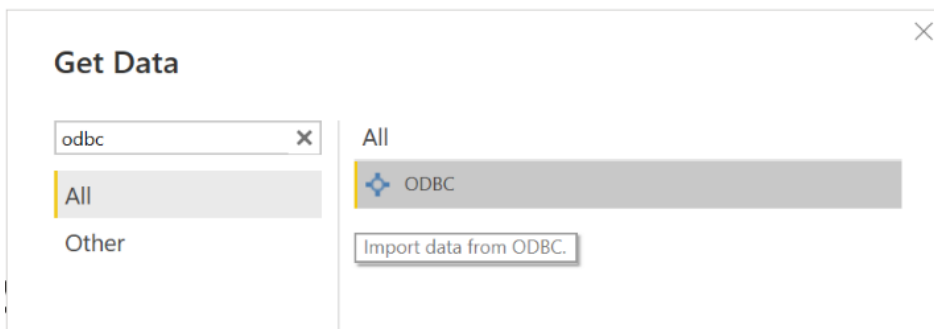
- To connect MongoDB with Power BI, we used the Simba ODBC driver which can be found here: https://www.magnitude.com/drivers/mongodb-odbc-jdbc?utm_source=google&utm_medium=search&utm_campaign=si-se-Drivers-RTU-2022&utm_content=121468984489&gclid=Cj0KCQjwmouZBhDSARIsALYcourDfv-jCtEEHgtGZ_mxVF_UTMvy6VtdL3MncMVR0Dsc-n7UOJBu6KYaAp0yEALw_wcB
- You will receive and email for simba ODBC license. Download the license and paste it under C:\Program Files\Simba MongoDB ODBC Driver\lib directory. The license file should look like this,



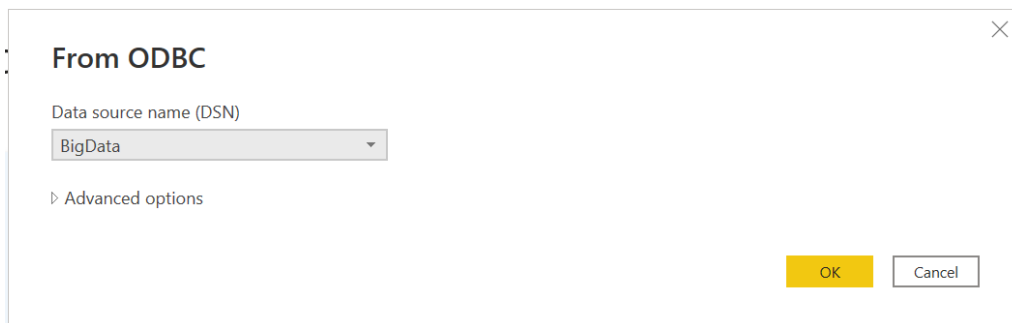
- After downloading and Installing Simba ODBC create a system DSN like the figure given below,



- Now open Power BI and select ODBC then click connect



- Select the DSN that you have created, in my case it is BigData



- Now you can find all the datasets in Power BI which were in MongoDB

