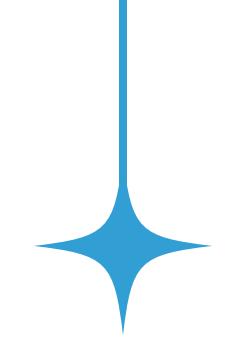


Muhammad Talha
BSCS | FAST NUCES

Power BI

For Titanic Dataset Exploration





Introduction

Power BI is a Microsoft's self-service BI platform for both on both premise and cloud-based data.

It is a business analytics tool used mainly for interactive visualizations i.e. to display charts and dashboards and create reports. It can also be used to clean up data and write DAX (Data Analysis eXpressions) formulas.



Power BI Desktop

Data Gateway

Paginated Report Builder

Power BI for Mobile

Analyze in Excel updates

Download/Install configuration

You can download from this link

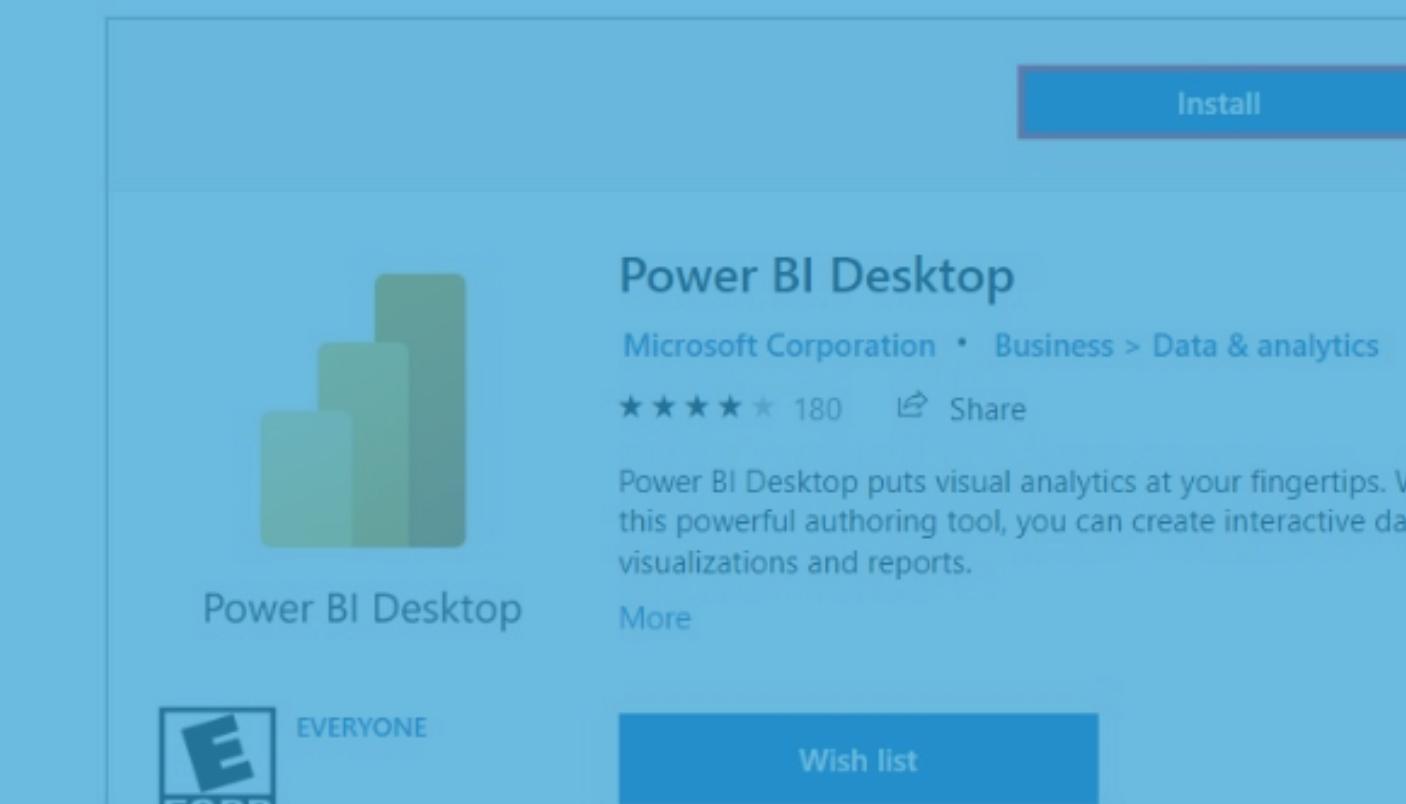
<https://powerbi.microsoft.com/en-us/desktop/>

Installation is simple as you do for other software's

No need to login

- Go to the Power BI Desktop product page[↗], then select Download Free.

2. After you've landed on the Power BI Desktop page in the Microsoft Store, select Install.



Dataset --- Titanic

- Rich data set with diverse variables
- Ideal platform for exploring relationships and identifying patterns
- Opportunity to practice DAX and Python for data transformation and analysis

Data Summary (i.e. columns):

survival (0= No, 1 =Yes)

pclass/ticket class (1=1st, 2=2nd, 3=3rd)

sex (female, male)

age (in years)

sibsp (number of siblings/spouses aboard)

parch (number of parents/children aboard)

ticket number

fare

cabin number

embarked (port of embarkation; C = Cherbourg, Q= Queenstown, S = Southampton)

Transformation using PQE

R W M J Y
O D G T
U C I B Z V
E L S A F K
X N H A P

R W M J Y
O D G T
U C I B Z V
E L S A F K
X N H A P

DATA



The screenshot shows the Microsoft Power BI Data Editor interface. On the left, the 'Queries [5]' pane lists 'titanic' as the active query. The main area displays a preview of the 'titanic' table with columns: Sex, Age, SibSp, and Parch. A formula bar at the top contains the expression: `= Table.ReplaceValue(#"Filtered Rows4", null,0,Replacer.ReplaceValue,{"Age"})`. To the right, a 'Query Settings' pane shows the properties and applied steps for the 'titanic' query. The 'APPLIED STEPS' section lists various transformations such as 'Duplicated Column1', 'Removed Columns2', 'Filtered Rows2', etc.

The screenshot shows the Microsoft Power BI Data Editor interface. The 'titanic' query is active. The main area displays a preview of the 'titanic' table with columns: Pclass, Sex, Age, SibSp, and Parch. A formula bar at the top contains the expression: `= Table.AddColumn(#"Filtered Rows3", "Pclass_class", each if [Pclass] = 1 then "Upper" else if [Pclass] = 2 then "Middle" else "Lower")`. To the right, a 'Query Settings' pane shows the properties and applied steps for the 'titanic' query. The 'APPLIED STEPS' section lists various transformations such as 'Added Custom3', 'Filtered Rows8', 'Replaced Value4', etc.

Data Cleansing

Remove inconsistencies, errors, and duplicates from your data.

Data Shaping

Rearrange, restructure, and pivot your data to meet your analysis needs.

Data Enrichment

Add calculated columns, merge data sources, and enhance data quality.

Dax (calculated columns and measures)

DAX (Data Analysis Expressions) is a formula language used in Power BI to create calculated columns and measures. Calculated columns are new columns derived from existing columns, while measures are aggregations of data across multiple rows.

The screenshot shows the Power BI Data Editor interface. In the center, there is a table with columns: Parch, A/B/C Ticket, 1.2 Fare, A/B/C Cabin, and Surv. Above the table, a DAX formula is displayed:

```
= Table.AddColumn(#"Removed Columns8", "AgeCategory", each if [Age] = 0 then "N/A" else if [Age] <= 18 then "Teens" else if [Age] <= 30 then "Youth" else if [Age] <= 50 then "Adult" else "Aged")
```

The left sidebar shows the 'Queries' list with 'titanic' selected. The right sidebar shows the 'APPLIED STEPS' list, which includes 'Added Custom3' (highlighted with a blue arrow).

Here are some examples of DAX expressions for the Titanic dataset:

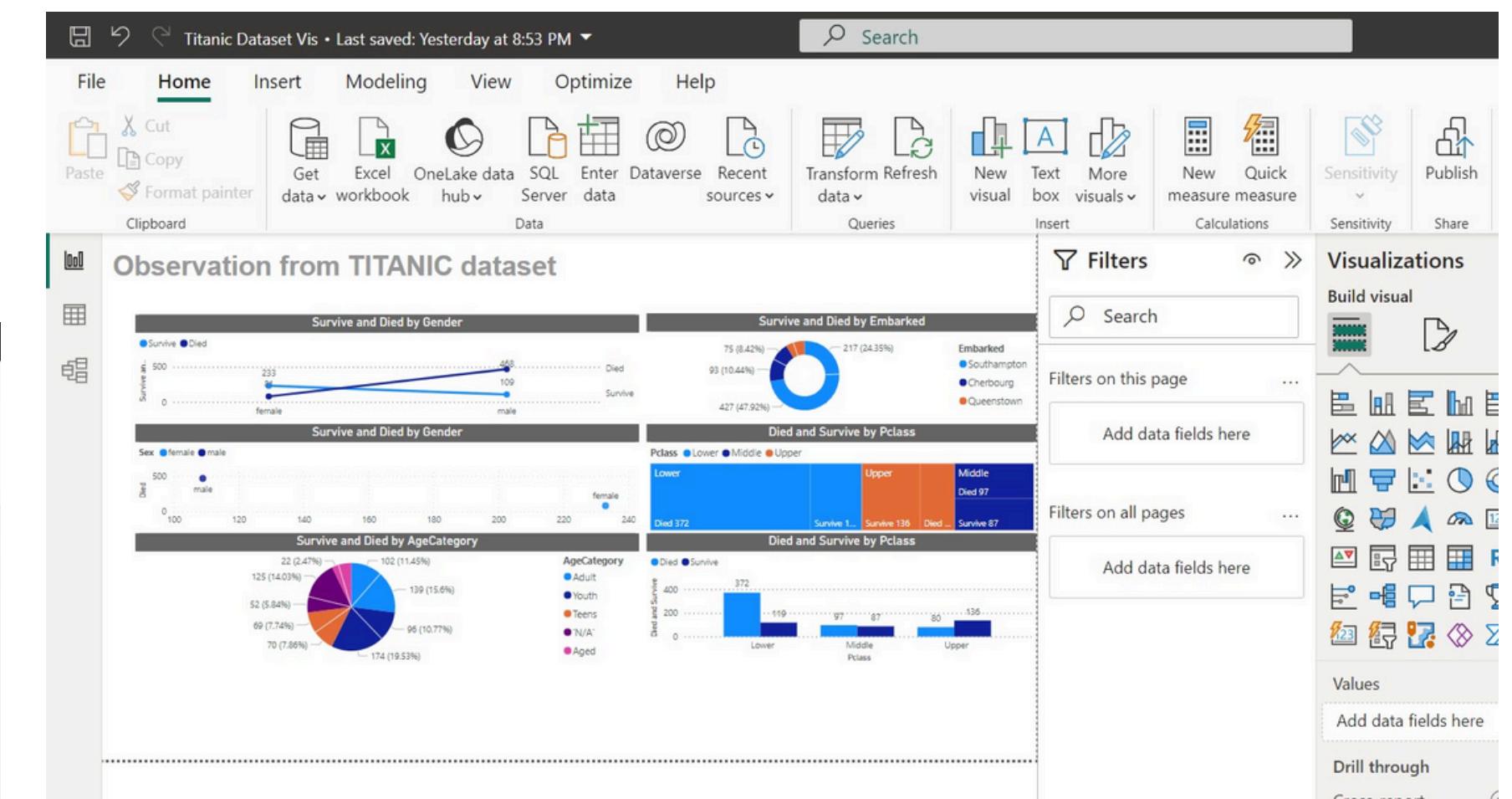
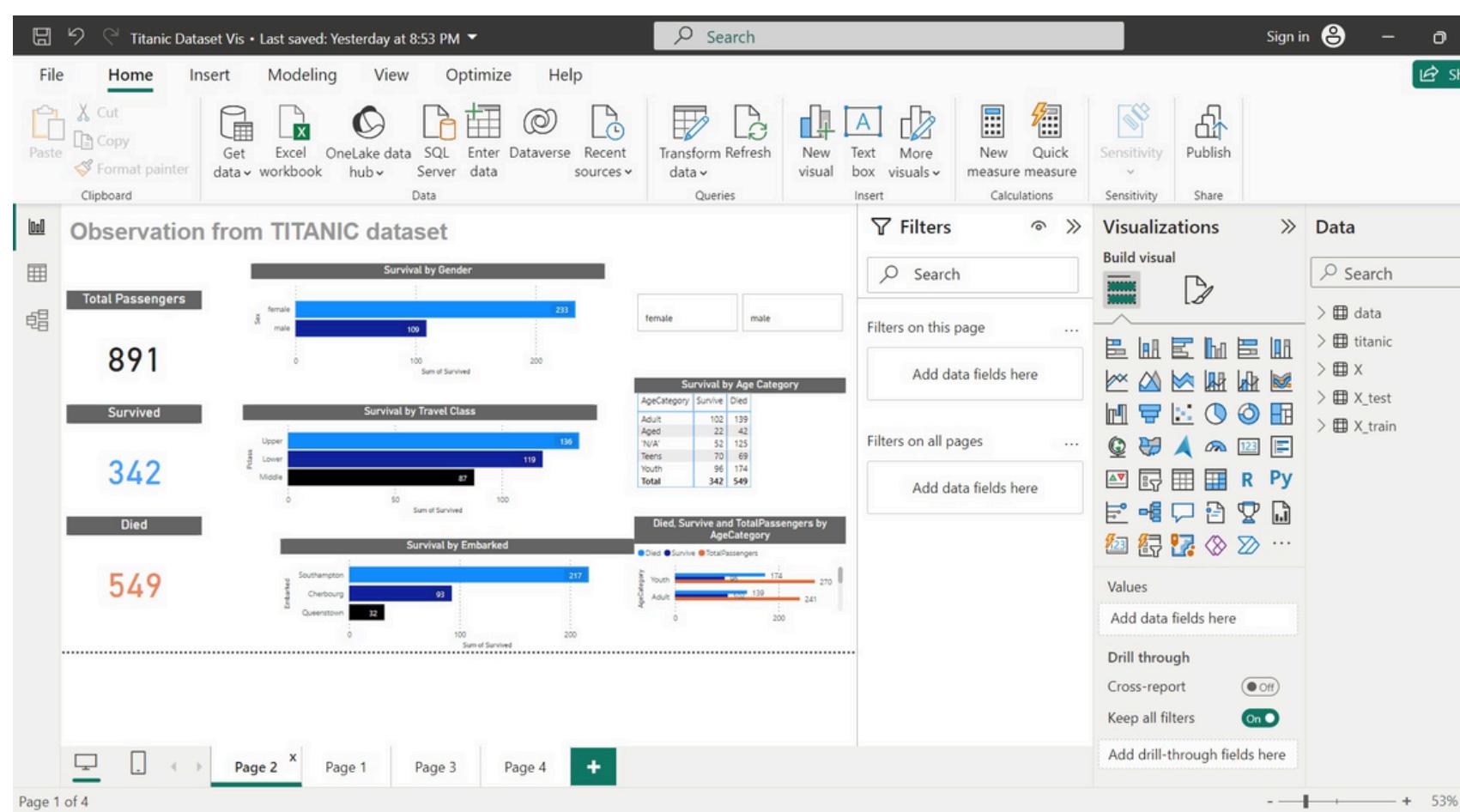
The screenshot shows the Power BI Model view. At the top, under 'Measure tools', a measure named 'Survive' is defined with the expression: `1 Survive = CALCULATE(COUNTROWS('titanic'), 'titanic'[Survived] = 1)`. The 'Data' pane on the right lists various columns and their data types, including 'Survived' (Boolean), 'Pclass' (Text), and 'Fare' (Number). A blue arrow points from the 'Measure tools' section towards the 'Data' pane.

Visualization with Power BI Charts



I didn't do much, I don't do much lol so I just;

- Inserted a text box and placed in my title at the top of the page.
- Created 3 different cards to display: a) Total Passengers b) Those who survived c) Those who died.
- Used bar charts to display the number of passengers who survived by a) Gender b) Point of embarkation c) Travel class
- Used a metric/table visual to display the number of passengers who survived by age category.
- Created a slicer which appeared at the top of the page to provide more granular information between Male or Female.

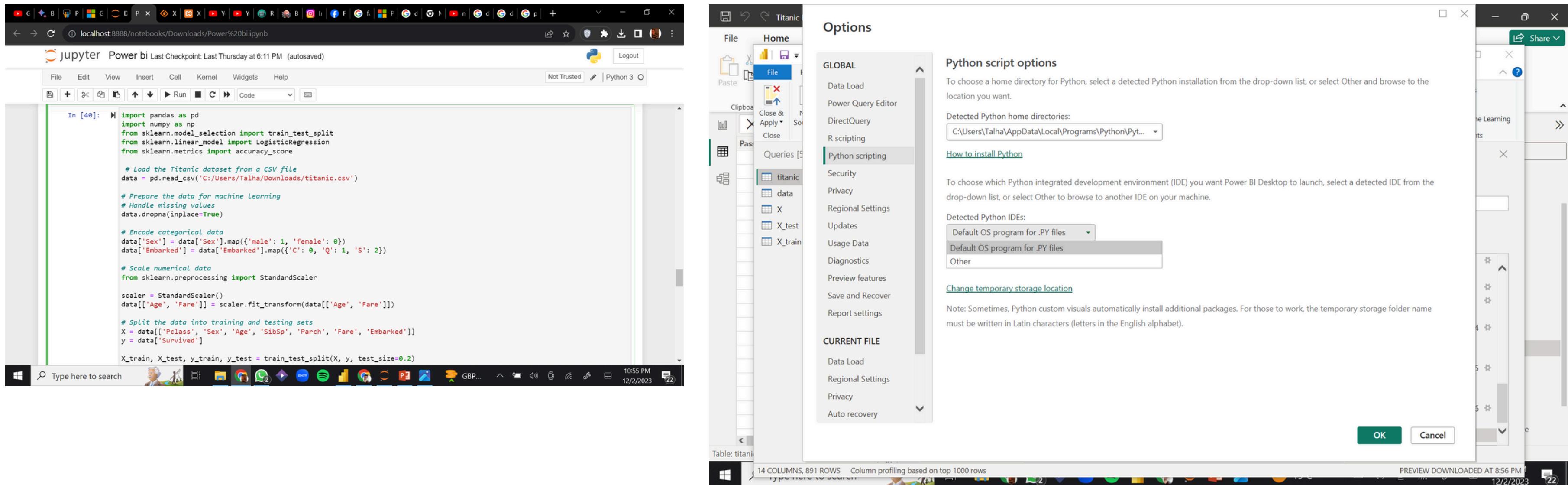


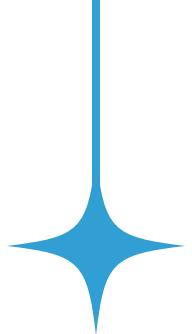
ML INTEGRATION

- Machine learning is a powerful tool that can be used to gain insights from data and make predictions.
- Power BI is a business intelligence tool that can be used to visualize and analyze data.
- By integrating Python into Power BI, you can extend the capabilities of Power BI to include ML.



I didn't do much, I don't do much lol so I just;





I didn't do much, I don't do much lol so I just;

Titanic Dataset Vis • Last saved: Yesterday at 8:53 PM

File Home Help Table tools Measure tools

Paste Cut Get data SQL Enter Transform Manage New Quick New New Manage View

Clipboard

1 Survive = CALCULATE(COUNTIF(PassengerId, Survived))

PassengerId	Name
6	Moran, Mr. James
20	Masselmani, Mrs. Fatima
27	Emir, Mr. Farred Chehab
29	O'Dwyer, Miss. Ellen 'Nellie'
30	Todoroff, Mr. Lailo
33	Glynn, Miss. Mary Agatha
37	Mamee, Mr. Hanna
43	Kraeff, Mr. Theodor
46	Rogers, Mr. William John
48	O'Driscoll, Miss. Bridget
77	Staneff, Mr. Ivan
78	Moutal, Mr. Rahamin Haim
83	McDermott, Miss. Bridget D
88	Slocovski, Mr. Selman Franc
96	Shorney, Mr. Charles Joseph
102	Petroff, Mr. Pastcho ("Pentch")
108	Moss, Mr. Albert Johan
122	Moore, Mr. Leonard Charles
127	McMahon, Mr. Martin
155	Olsen, Mr. Ole Martin
160	Emiliano, Mr. Mila

Python script

```
# Encode categorical data
data['Sex'] = data['Sex'].map({'male': 1, 'female': 0})
data['Embarked'] = data['Embarked'].map({'C': 0, 'Q': 1, 'S': 2})

# Scale numerical data
from sklearn.preprocessing import StandardScaler

scaler = StandardScaler()
data[['Age', 'Fare']] = scaler.fit_transform(data[['Age', 'Fare']])

# Split the data into training and testing sets
x = data[['Pclass', 'Sex', 'Age', 'SibSp', 'Parch', 'Fare', 'Embarked']]
y = data['Survived']
```

The script will run with the following Python installation
C:\USERS\TALHA\APPDATA\LOCAL\PROGRAMS\PYTHON\PYTHON312.

To configure your settings and change which Python installation you want to run, go to Options and settings.

OK Cancel

Table: titanic (891 rows) Column: Survive (0 distinct values)

Search

Sign in

Share

Sensitivity

Clipboard

PassengerId

Pclass

Sex

SibSp

Survived

Survived_text

Ticket

TotalPassengers

TotalTransportationFare

10:52 PM 12/2/2023



**Thank You
For Your Time!**

