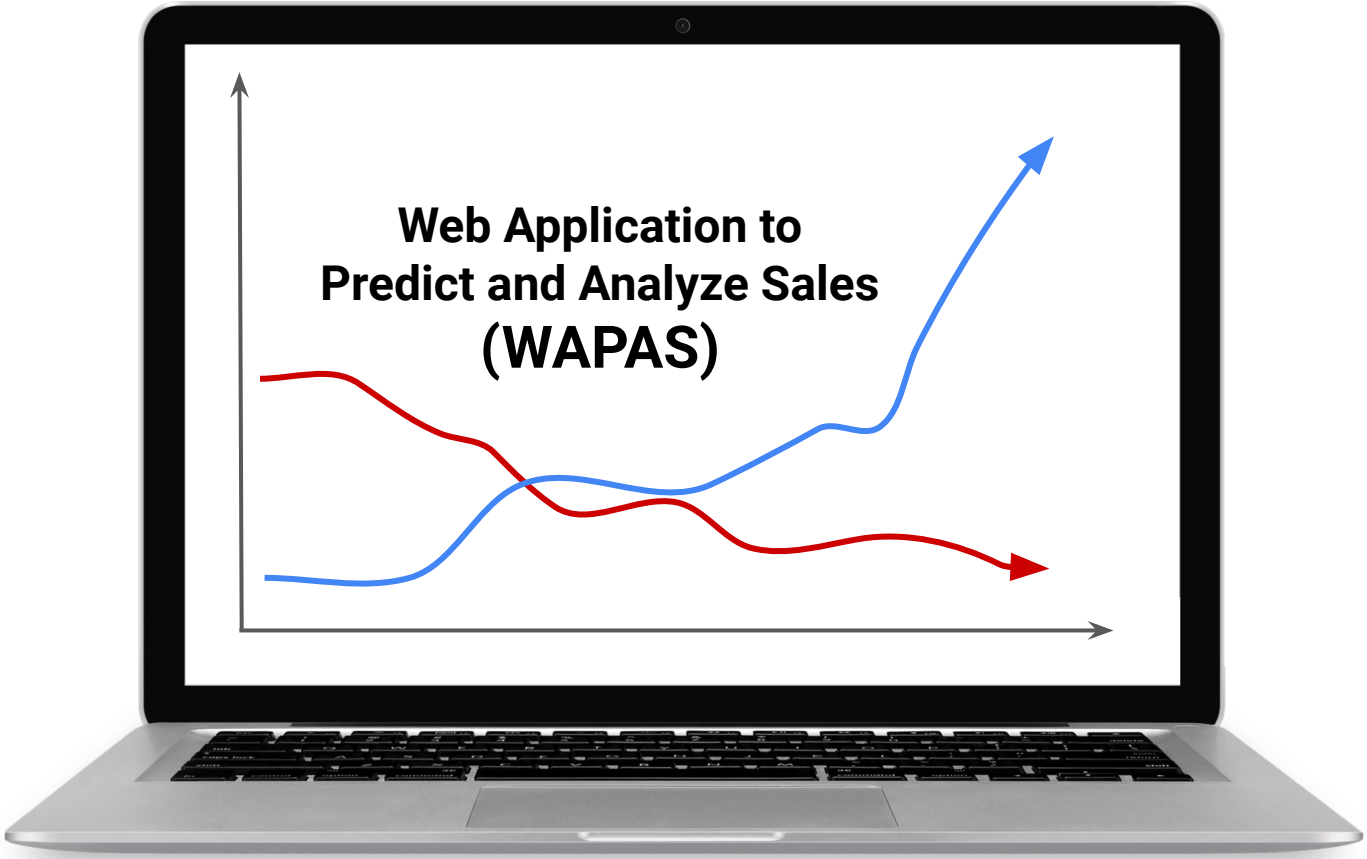
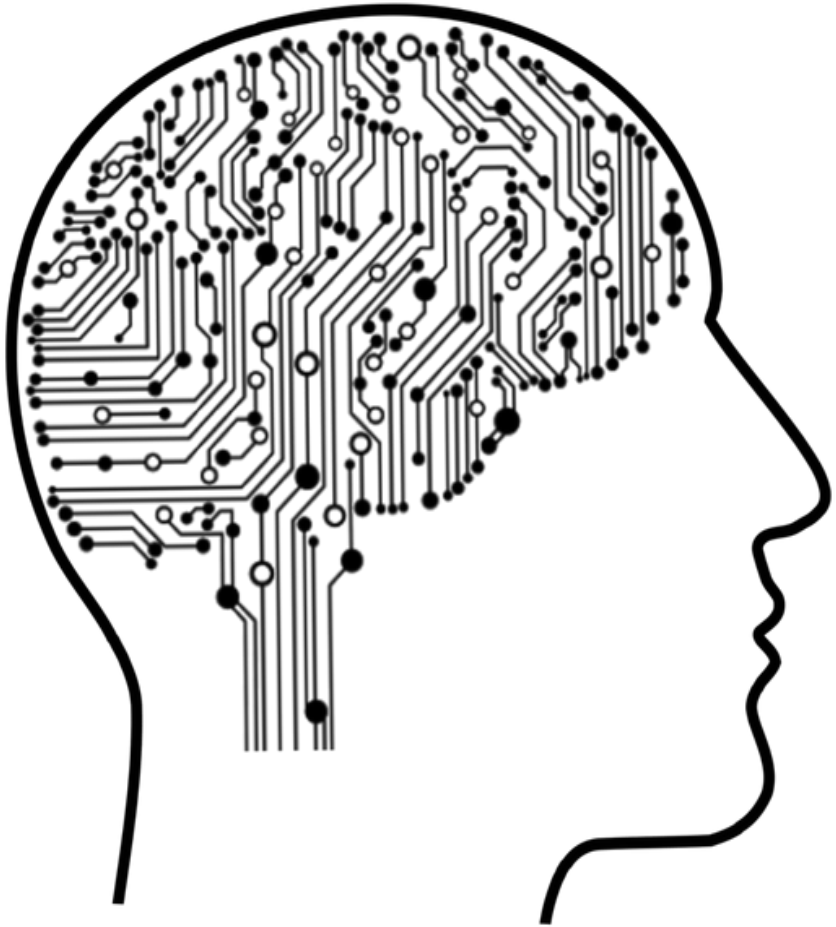

Web Application to
Predict and Analyze Sales
(WAPAS)



Project by,

Aritra Banerjee
Indranuj Ganguly
Ranadip Ghosh
Rounak Das
Sarat Sen

Under the guidance

of

Tuhina Sinha

What is Sales Analytics?

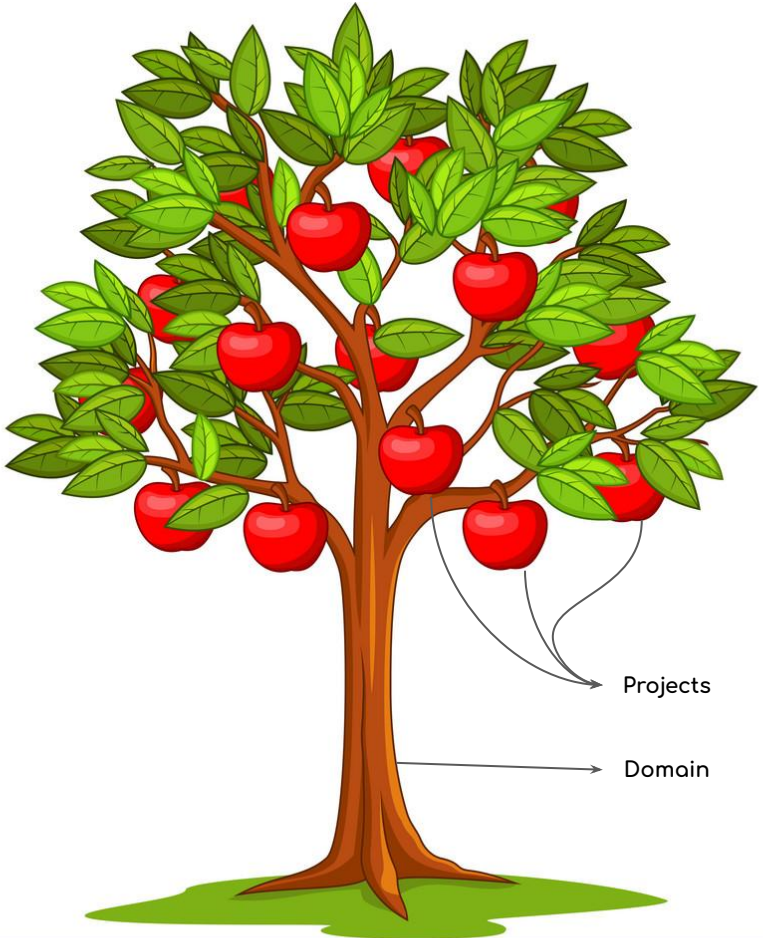
- ➔ Data-driven decision making
- ➔ Commercially relevant insights
- ➔ Business estimates
- ➔ Cost analysis
- ➔ Consumer behavioral understanding
- ➔ Brand perceptions

Revenue & profitability

Consumer sentiment to the brand

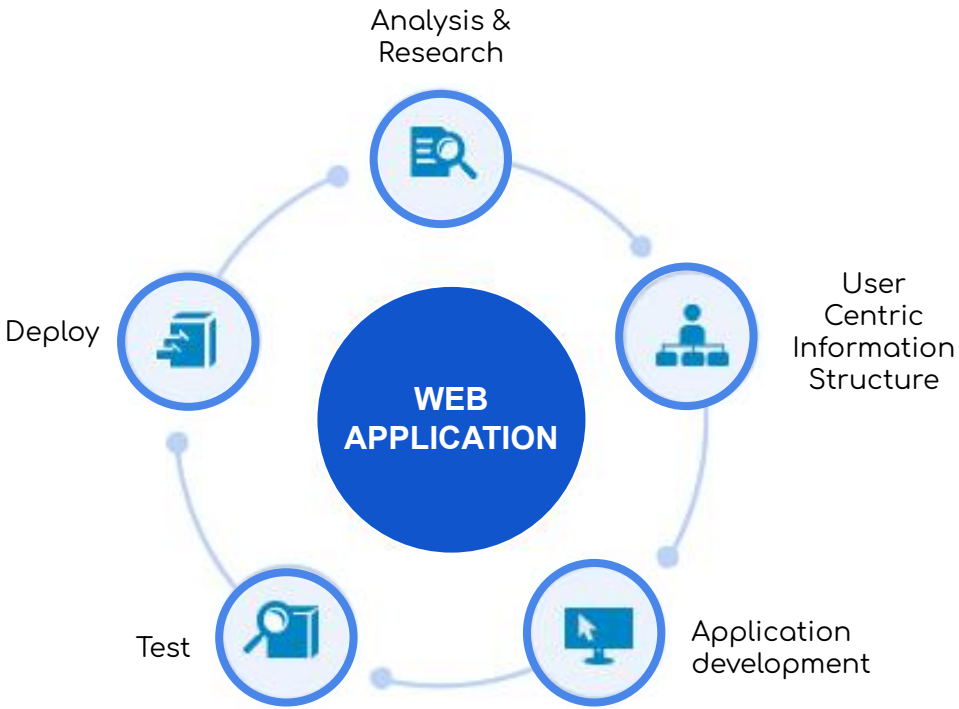
How WAPAS stands out?

- ❑ Simple, clean & intuitive UI
- ❑ Data visualization feature
- ❑ Reliable accuracy
- ❑ Efficient business outlook
- ❑ Secured by Unique user ID
- ❑ Reusability & maintenance

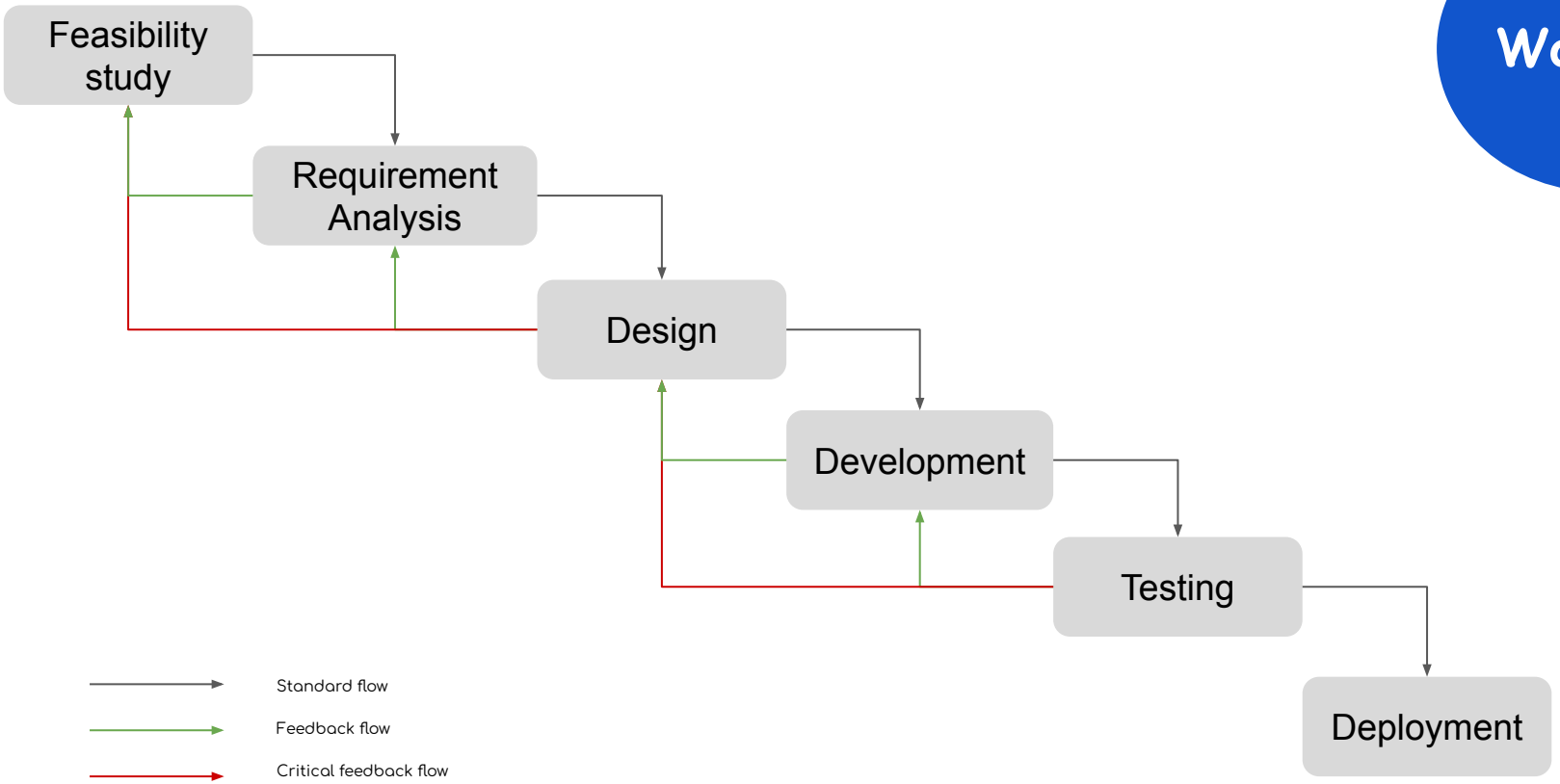


Why Web Application?

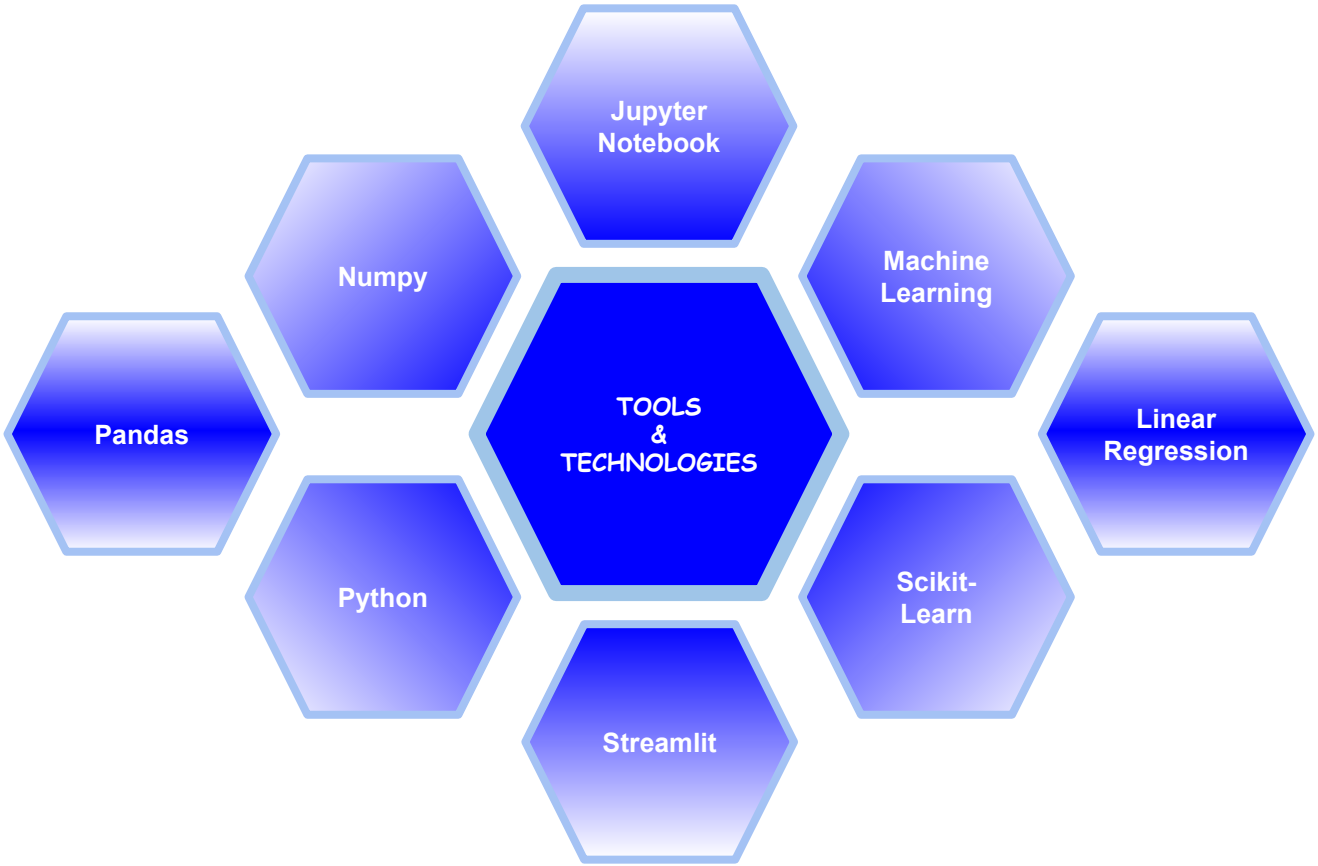
- User-friendly deployment
- Easily accessible
- Cross platform compatibility
- Comparatively less expensive
- Widespread manageability
- Simple authentication system



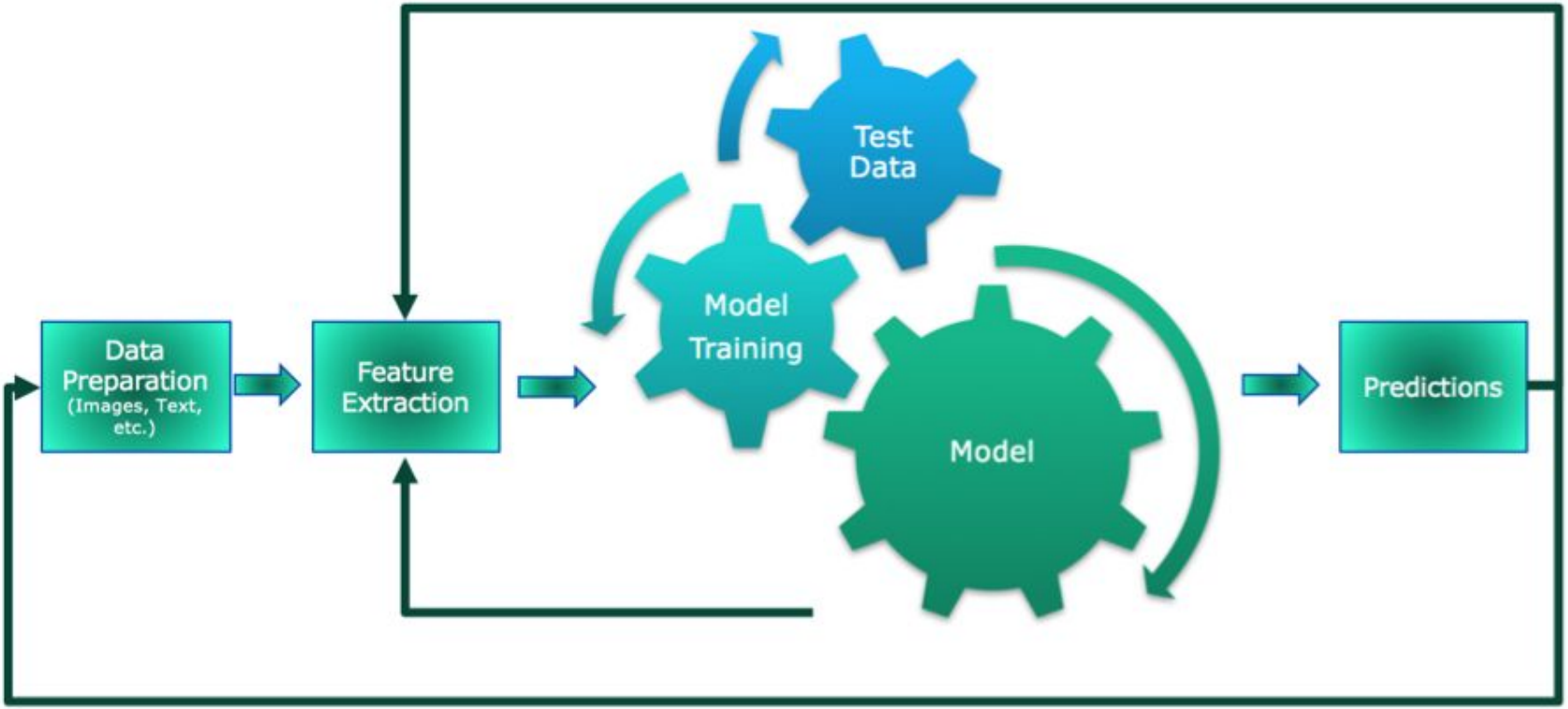
Workflow



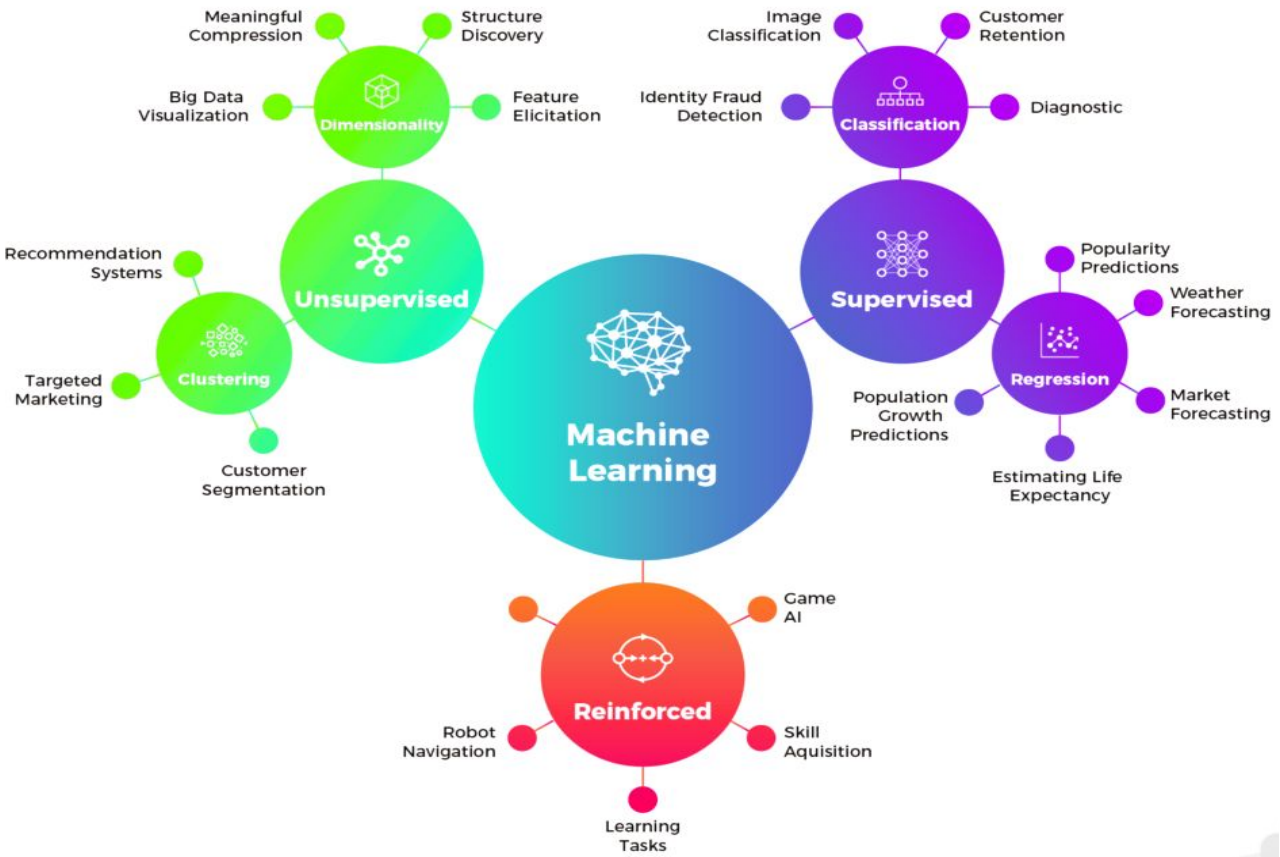
Royce's Waterfall model


Jupyter Notebook
Numpy
Machine Learning
Pandas
TOOLS & TECHNOLOGIES
Linear Regression
Python
Scikit-Learn
Streamlit

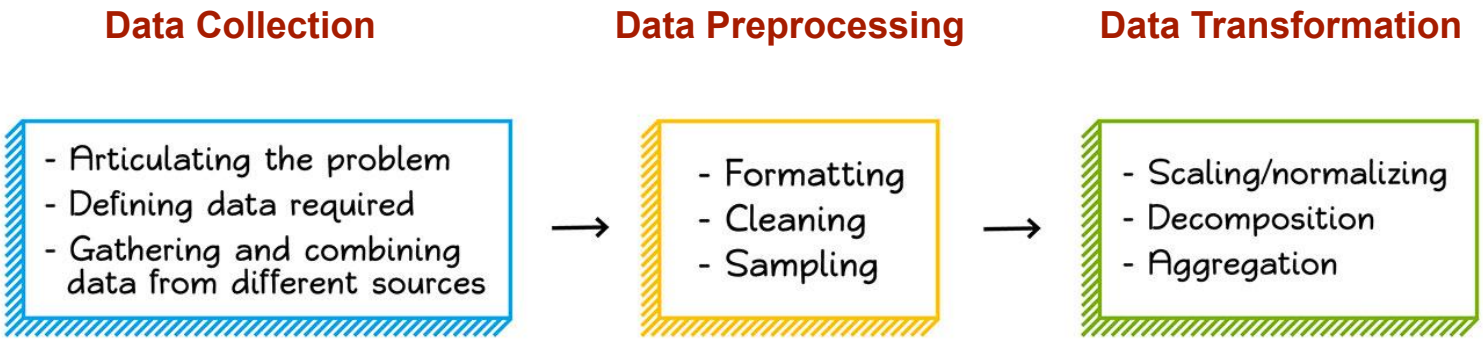
Standard ML pipeline



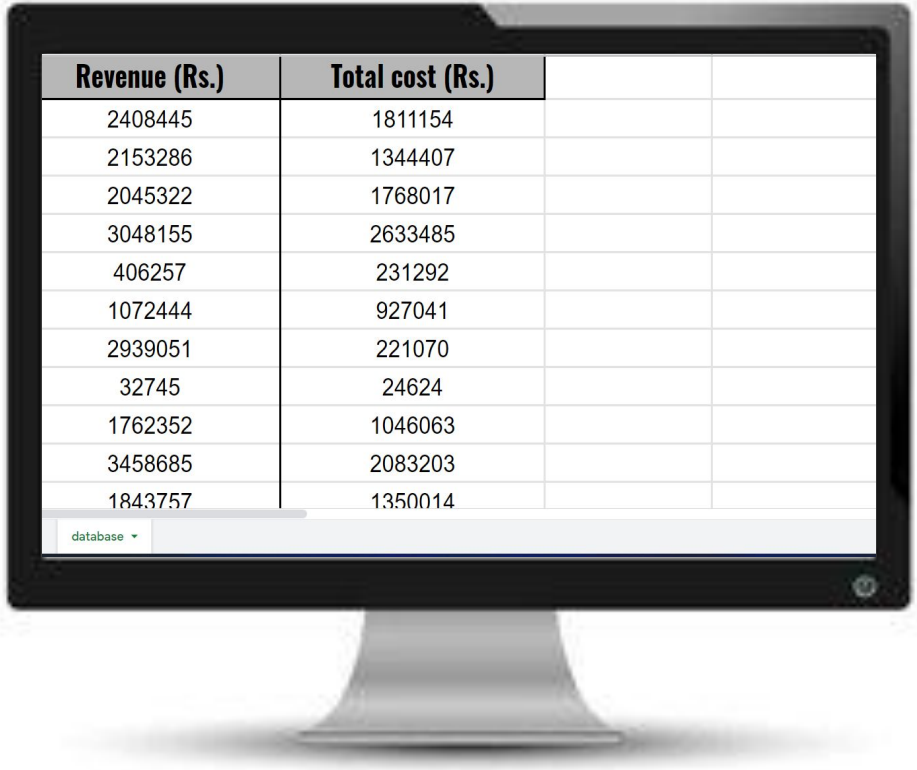
Different types of ML models



Data Preparation Process



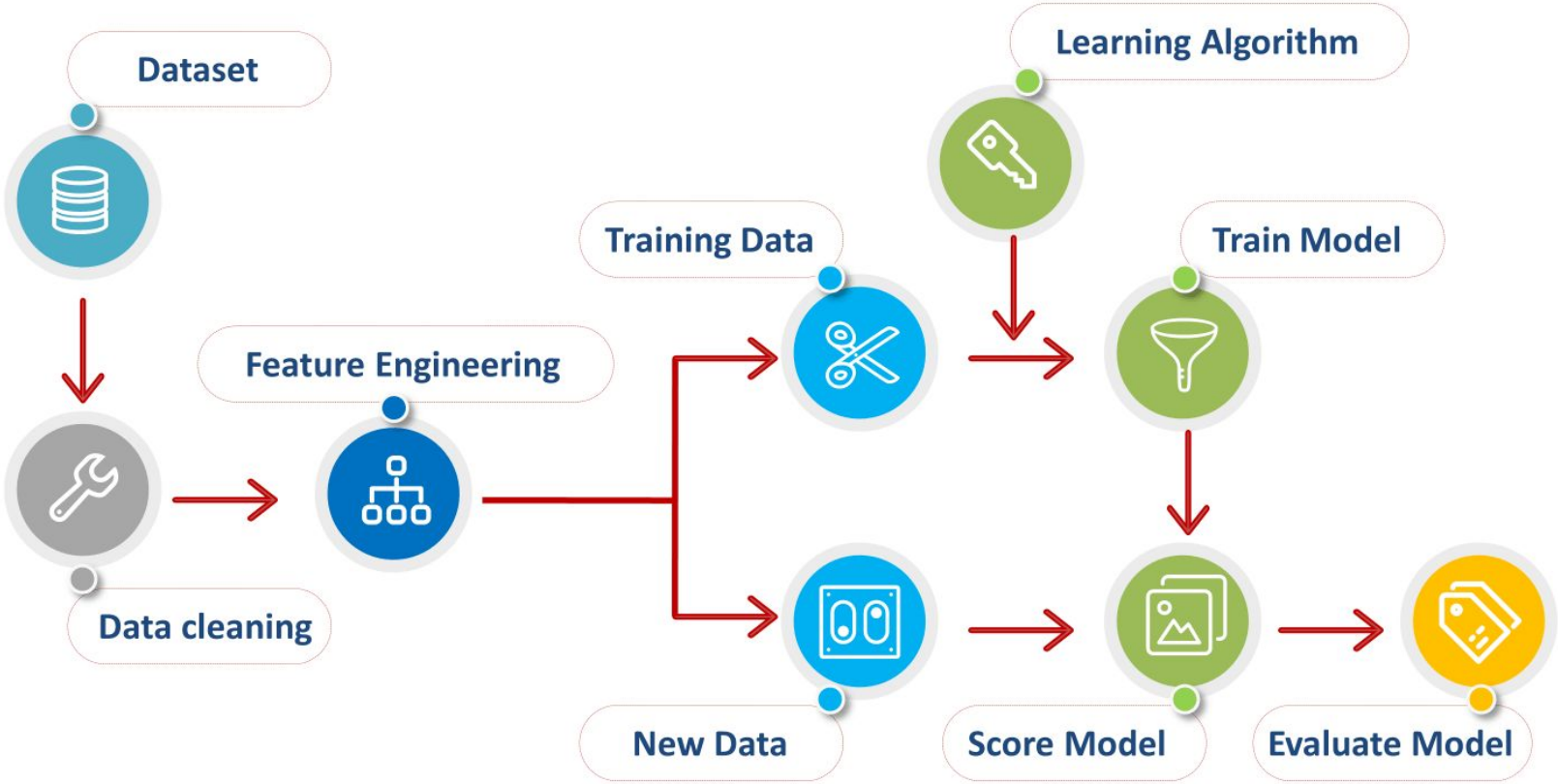
Sample data collection



Revenue (Rs.)	Total cost (Rs.)
2408445	1811154
2153286	1344407
2045322	1768017
3048155	2633485
406257	231292
1072444	927041
2939051	221070
32745	24624
1762352	1046063
3458685	2083203
1843757	1350014

database

Model Evaluation

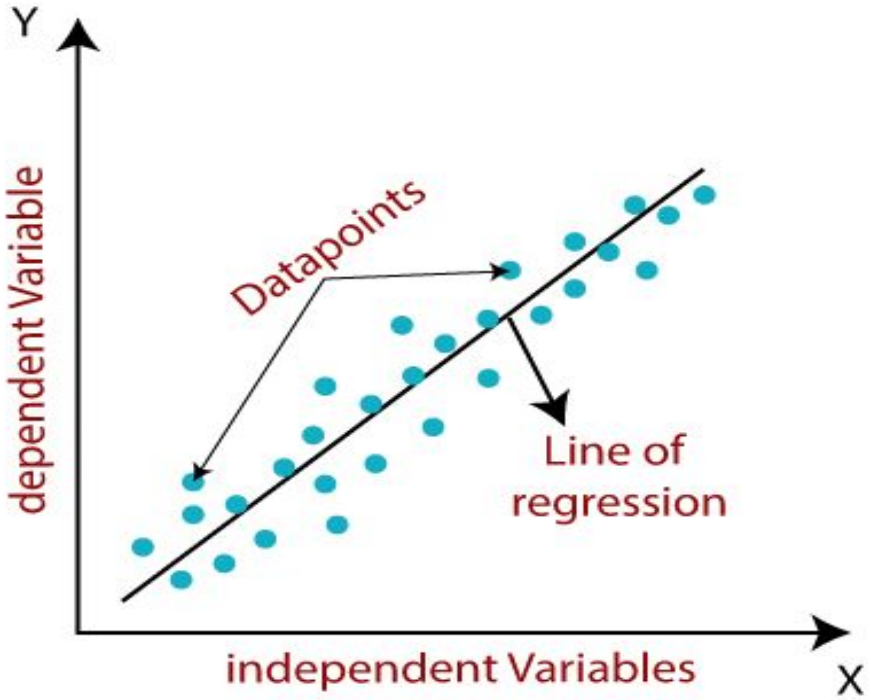


Linear Regression

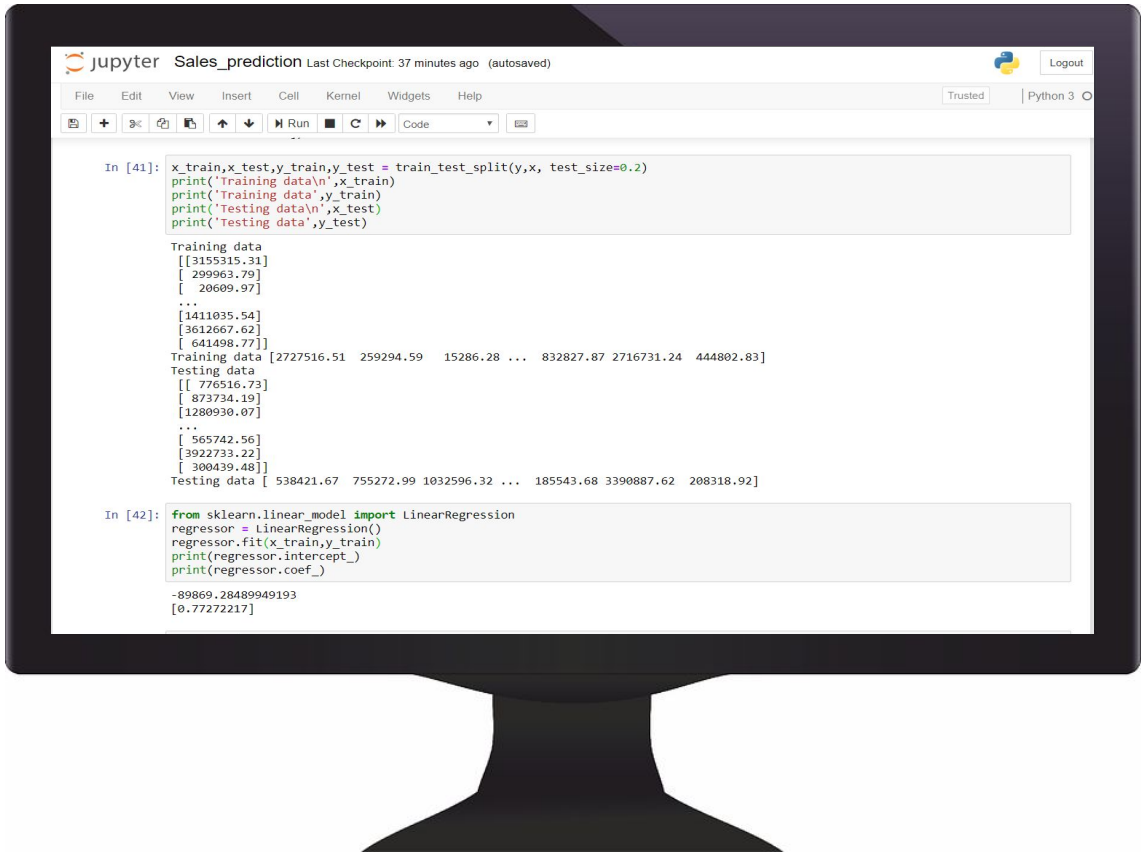
We have used ***Supervised Machine Learning*** to understand the sales record and predict its future scope.

To be more specific, we chose Linear Regression classification model for solving this problem.

$$Y = mX + C$$

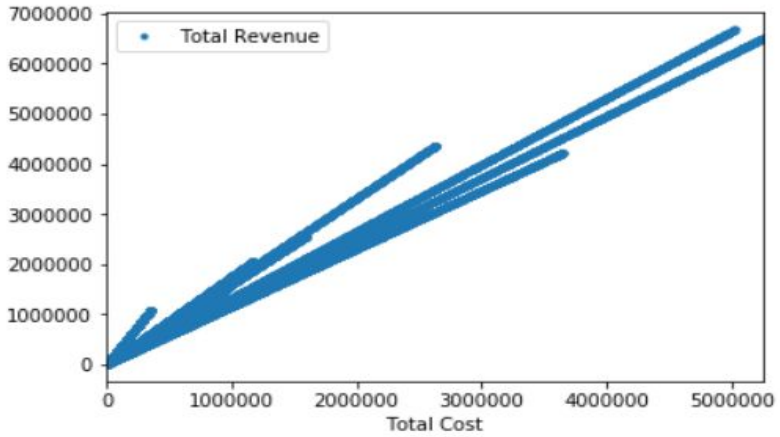


Notebook Snapshot



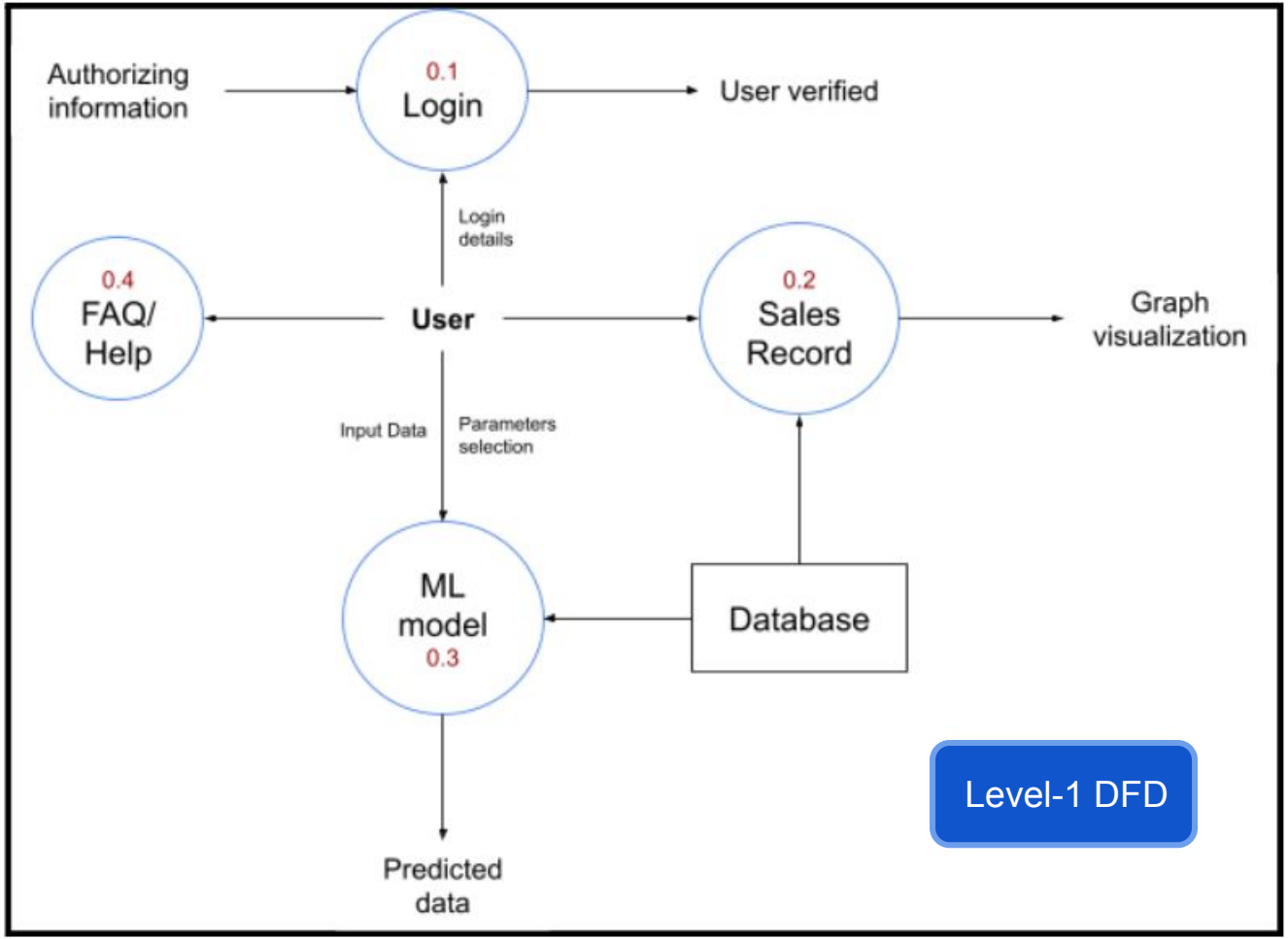
Visualization

```
In [39]: dataSet.plot(x='Total Cost',y='Total Revenue',style='.')
         plt.show()
```

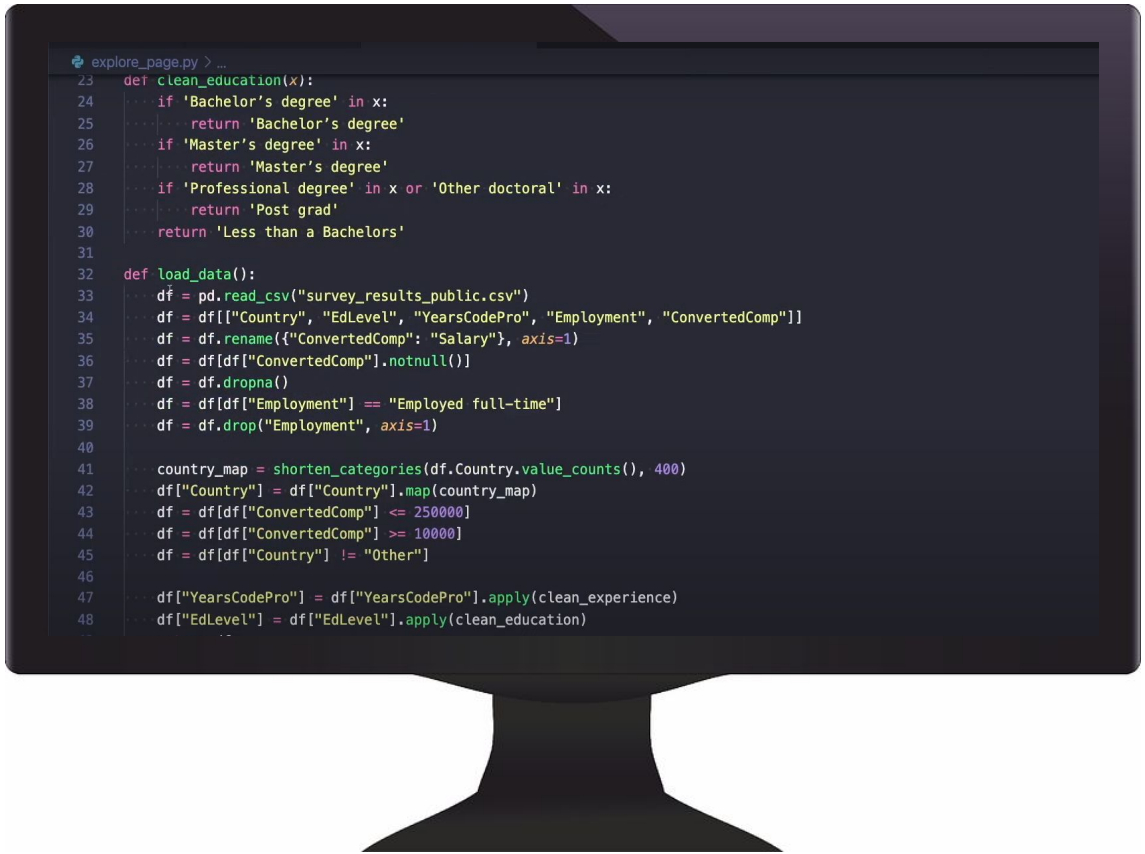


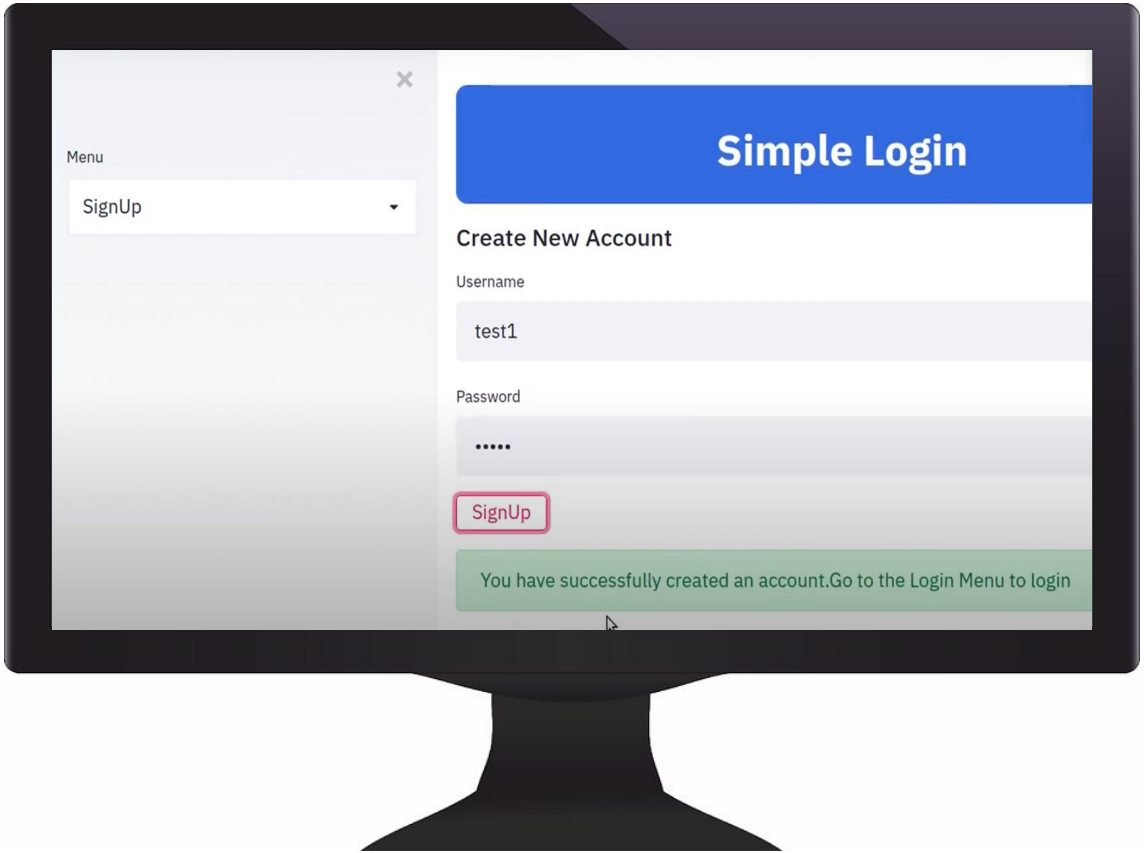
Web Application

- Login page
- Visual analysis of existing sales data
- Parameters-based prediction page
  - (Sales, revenue, profit & cost)
- Remark/note
- FAQ/Help page


Authorizing information
0.1
Login
User verified
Login details
0.4
FAQ/ Help
User
0.2
Sales Record
Graph visualization
Input Data
Parameters selection
ML model
0.3
Database
Predicted data
Level-1 DFD

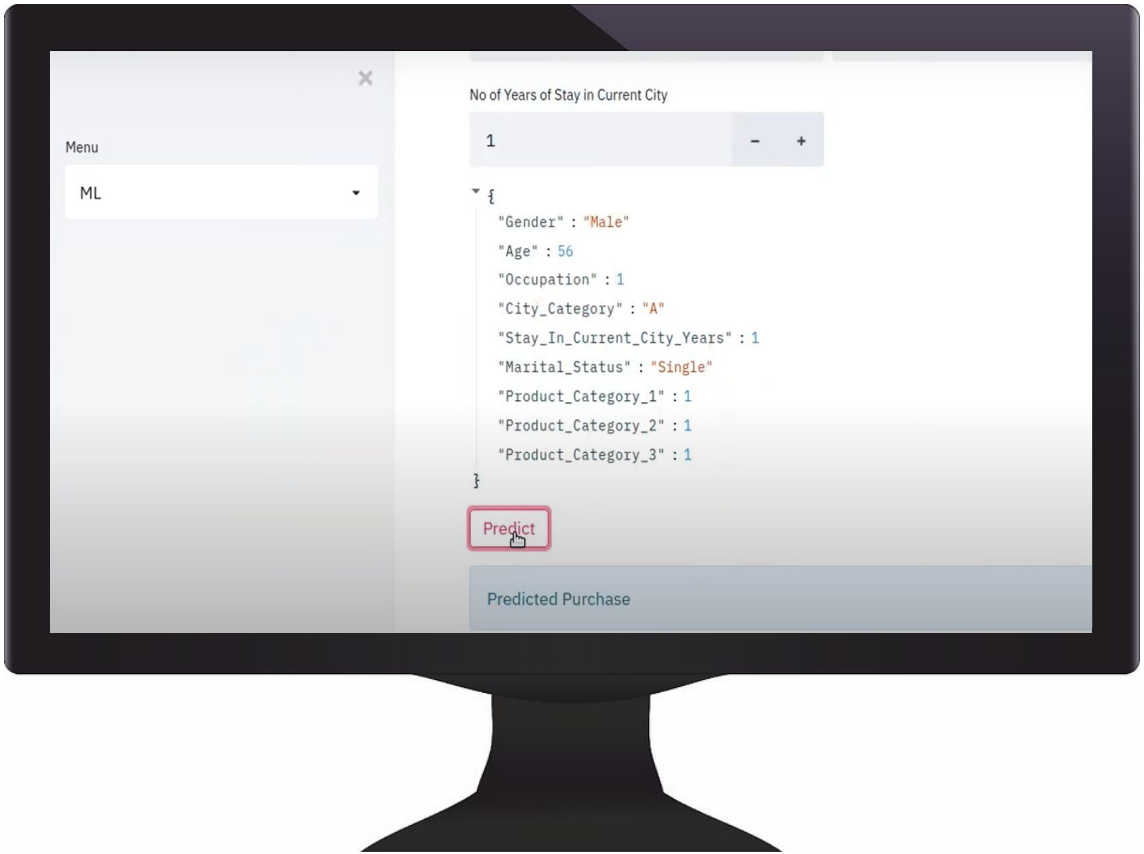
Sample code


explore_page.py > ...
def clean_education(x):
if 'Bachelor's degree' in x:
return 'Bachelor's degree'
if 'Master's degree' in x:
return 'Master's degree'
if 'Professional degree' in x or 'Other doctoral' in x:
return 'Post grad'
return 'Less than a Bachelors'
def load_data():
df = pd.read_csv("survey_results_public.csv")
df = df[["Country", "EdLevel", "YearsCodePro", "Employment", "ConvertedComp"]]
df = df.rename({"ConvertedComp": "Salary"}, axis=1)
df = df[df["ConvertedComp"].notnull()]
df = df.dropna()
df = df[df["Employment"] == "Employed full-time"]
df = df.drop("Employment", axis=1)
country_map = shorten_categories(df.Country.value_counts(), 400)
df["Country"] = df["Country"].map(country_map)
df = df[df["ConvertedComp"] <= 250000]
df = df[df["ConvertedComp"] >= 10000]
df = df[df["Country"] != "Other"]
df["YearsCodePro"] = df["YearsCodePro"].apply(clean_experience)
df["EdLevel"] = df["EdLevel"].apply(clean_education)


Menu
SignUp
Simple Login
Create New Account
Username
test1
Password
•••••
SignUp
You have successfully created an account.Go to the Login Menu to login

Login page

Sales analysis



Prediction page

Challenges faced

- Lack of clean data to directly work with might have slowed down our progress.
- The loss to value of information in a real scenario for sales is very high.
- Content based classification is just a part of the whole picture.
- Distinguish between sales fraud and actual sale.

Future scope

- ➔ Accuracy increment
- ➔ Data verification
- ➔ Additional model deployment
- ➔ Relevant data sources
- ➔ Range optimization
- ➔ Metrics adjustment
- ➔ Cross platform development
- ➔ Business outlook improvement
- ➔ Experienced market research


Thank
you

References

Dataset:
https://www.kaggle.com/aungpyaeap/supermarket-sales

Royce's Waterfall model:
https://en.wikipedia.org/wiki/Waterfall_model

Standard ML pipeline:
https://www.datanami.com/2018/09/05/how-to-build-a-better-machine-learning-pipeline/

Different types of Machine Learning:
https://www.rocketsource.co/blog/machine-learning-models/