

Data Science using Python

Agenda

What is data science

High level overview of data science

Practical Hands-on session using Python

Solve business problems in E-commerce

E-commerce dataset

Data Science using Python

What is data science?

Converting raw data into insights

Websites

Sensors

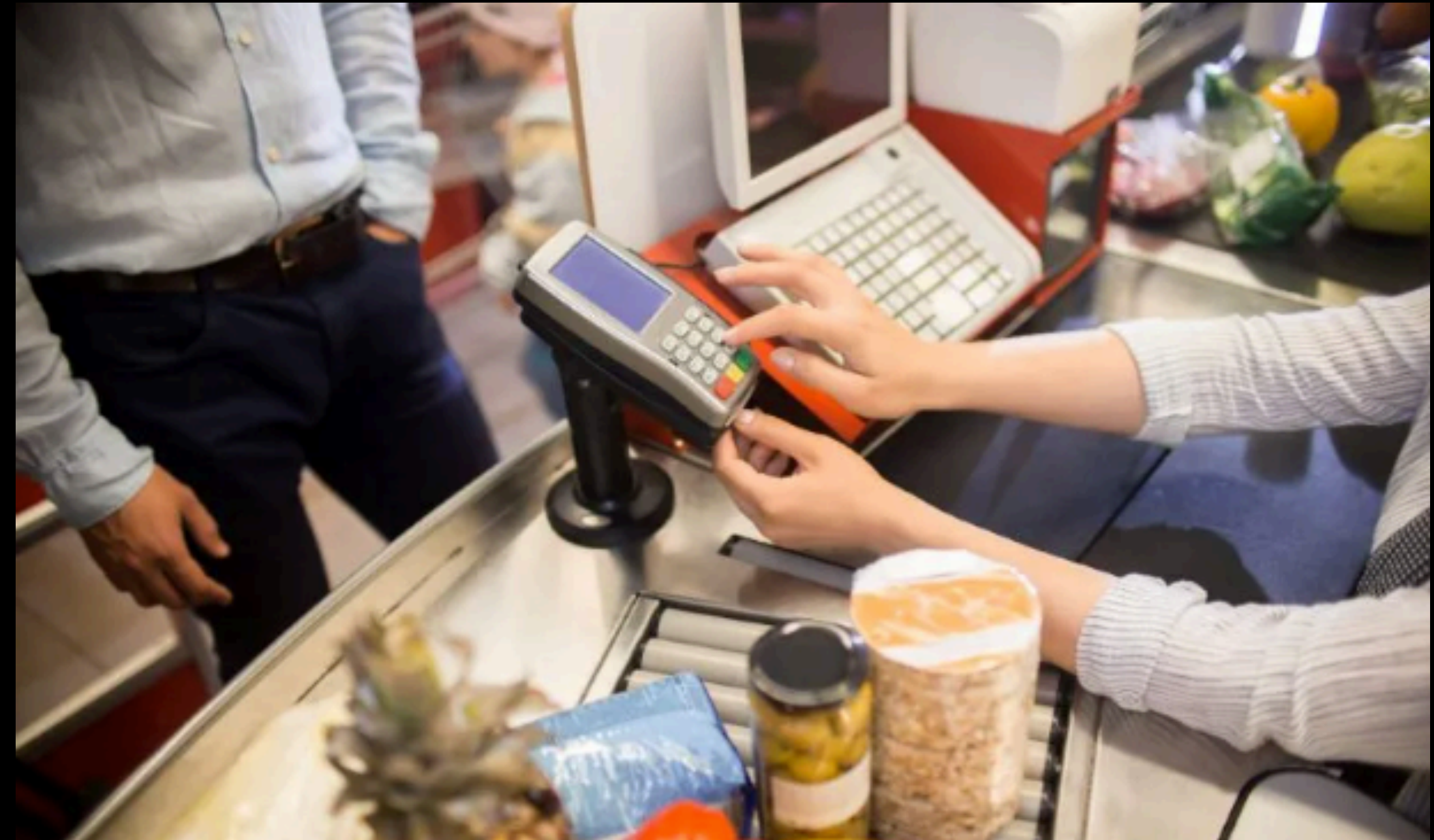
Surveys

Store the data

Get Insights

Specialized software
and skills

Stakeholders



QARA Wood Laminated Study Table, Computer Table for Home, Office Table, Desktop, Laptop Table, Office Desk (White)

Visit the QARA Store
4.5 ★★★★★ 396 ratings
400+ bought in past month

-63% ₹3,149
M.R.P.: ₹8,699

Inclusive of all taxes
EMI starts at ₹153. No Cost EMI available EMI options ▾

Coupon: Apply ₹50 coupon [Terms](#)

Offers

Bank Offer Upto ₹1,750.00 discount on select Credit Cards, HDF... 25 offers >	No Cost EMI Upto ₹141.81 EMI interest savings on Amazon Pay ICICI... 1 offer >	Partner Offers Get GST invoice and save up to 28% on business purchase... 1 offer >
--	---	--

Free Delivery Cash/Pay on Delivery 10 days Replacement Top Brand Installation available Amazon Delivered

Data Science using Python

E-Commerce

Suppose you see a dip of 15% in sales

You need to know the reason behind this

Suppose the stakeholders need to know how much to stock?

You need to predict sales for the next week

Typical steps to analyse

Check website data

Load this into your system (typically SQL)

Clean the data (typically Python)

Find patterns (typically plot with matplotlib/seaborn)

Predictions using various attributes (“features”) of the data

Data Science using Python

Correlation

Croma - Launch AC

What leads to rise in sales?

Temperature

Season

Cost

Power savings (4 star, 5 star)

Ice cream sales

Data Science using Python

Python - Hands on

Basic intro to Python

Google Colab to write code <https://colab.research.google.com>

Libraries: pandas and seaborn

Dataset - ecommerce

<https://community.tableau.com/s/contentdocument/0694T000001GnpUQAS>

Order Date, Customer ID, Segment, Country, Region,
Product id, Quantity, Discount, Profit

Data Science using Python

Problem statement - Reason for drop in sales

State level profits

Top 3 states having loss

Category level

Top categories with high loss

Find correlation that leads to higher profit or sales

Data Science using Python

Deep dive

Upload the data to colab

Import the relevant libraries

Pandas and Seaborn

Read the File

Use the “read_excel” method from pandas

Assign to python variable “df”

Observe a small portion of the data

`df.head, df.tail`

Data Science using Python

State wise profit

Group the data on the basis of State

Total profit for each state

Total sales for each state

State with maximum loss

Sort the values

Data Science using Python

Top loss making categories

Group the data on the basis of State

Total profit for each state

How many orders for each category

Data Science using Python

Correlations in python

```
df.corr()
```

```
sns.heatmap()
```

Resources

Scaler Topics: www.scaler.com/topics

Hacker Rank: hackerrank.com

Corey Schafer: <https://www.youtube.com/@coreyms>

Jake Vanderplas: <https://jakevdp.github.io/PythonDataScienceHandbook/>

ThinkStats: <https://greenteapress.com/wp/think-stats-2e/>

Sebastian Rashcka: <https://github.com/rasbt/machine-learning-book>

Software Engineering for Data Scientists

- <https://github.com/catherinenelson1/SEforDS>