

# Data Analysis of Amazon Reviews



Anirudh Rao  
Great Learning

# Agenda

## Key Takeaways

- **Introduction** to Data Analytics
- **Types** of Data Analysis
- **Python** programming for Data Analysis
- Introduction to **Web Scrapping**
- **Hands-on Demo:** Amazon Review Analysis



# Introduction to Data Analytics

# Introduction to Data Analytics

## What is Data Analytics?

- Process of collecting data, evaluating it and converting it into useful information.
- Examples include sales data, logistics analysis and market research.
- Involves working towards accuracy of data.
- Data presentation is a key part of Data Analytics as well.



# Introduction to Data Analytics

## Who is a Data Analyst?

- An expert who can work with multiple teams such as IT, management, data science teams to understand operational goals.
- Responsible for gathering data, cleaning it and providing key business insights and hidden trends.



# Introduction to Data Analytics

## Task overview of a Data Analyst!

- Designing database systems
- Preprocessing the data
- Applying Machine Learning algorithms
- Creating beautiful looking dashboards



# How to become a Data Analyst?

# Skills required to become a Data Analyst

Data Analysis is a **vast** domain!

- Programming languages (Python, R or SAS)
- Creative thinking
- Communication skills
- SQL and database query languages
- Machine learning

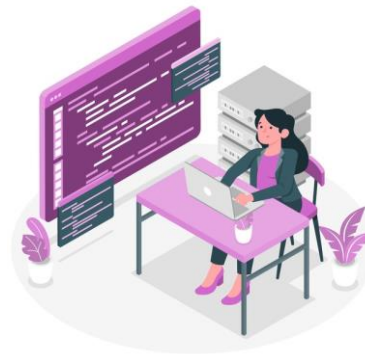




# Important tools used by Data Analysts

Usage of tools depend on the domain!

- Jupyter Notebook
- Tableau
- Apache Spark
- Microsoft Excel
- Splunk



# Data Analyst



## Popular Data Analyst designations



**Associate Data**  
Analyst



**Senior Data**  
Analyst



**Lead Data**  
Analyst



**Product**  
Manager

# Data Analysis Types



# Data Analysis: Types

**Analytics:** The varieties!

- Descriptive analytics
- Predictive analytics
- Prescriptive analytics



# Descriptive Analytics

“What happened here?”

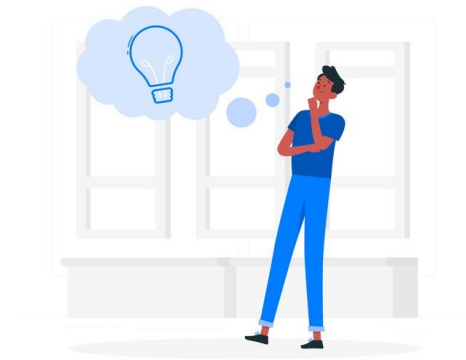
- Why did the sales decrease this much?
- How did this product become popular?
- What are the reasons for the sale spike?



# Predictive Analytics

“What is likely to happen?”

- What product should be stocked up for the holidays?
- What returns do we get on marketing campaigns?
- What will be the trending product next summer?



# Prescriptive Analytics

“What action should be taken?”

- How to bridge the supply-demand gap?
- How to develop product based on customer feedback?
- How to provide better products and service?





# Getting started with programming



# Getting started with programming

Programming is an **important** aspect of Data Analytics!

- Python is the popular choice for Data Analytics.
- Step-by-step learning is important
- **Prime tip:** Take it slow and steady
- Try to learn using the practical approach



# Python programming

World's most popular programming language!

- High-level programming features
- Object-oriented language
- Simple to use and versatile
- LARGE community
- Key role in Data Science and AI related domains.



# Where is Python Used?

You're using a product of Python right now!

- Watching videos on YouTube
- Booking a cab or a flight
- Opening an internet browser
- Listening to music on streaming services
- Video streaming services





# Popularity of Python

# Why is Python this popular?

Very **important** reasons you know!

- Readable and Maintainable Code
- Multiple Programming Paradigms
- Thousands of libraries
- Open Source Frameworks and Tools
- Simplified Software Development
- Test-Driven Development



# Companies that use Python

Thousands of companies use Python!

- Google
- Facebook
- Instagram
- Netflix
- Quora
- Spotify



# Skills Needed to Learn Python!

# Skills Needed to Learn Python

## Important Requirements!

- Data Oriented Approach
- Out of the box thinking
- Logic Building Skills
- Ability to scale up quick





# Introduction to Web Scraping

# Introduction to Web Scraping



## What is Web Scraping?

- The information is collected from the websites according to the need and then exported those data into a proper format.
- Web scraping can be done manually as well.
- But normally preferred to use automated tools for the web scraping.

# Why is Web Scraping needed?



# Is Web Scraping legal?

# Is Web Scraping legal?

Yes, it is legal!

- Yes, Web scraping is legal.
- You are allowed to scrape anything which is publicly available.
- But, if you are sending multiple requests in a very short span of time to Amazon then your real IP address will be banned.



# Python libraries for Web Scraping

# Python libraries: Web Scraping

Many popular libraries to choose from!

- Requests
- BeautifulSoup 4
- Lxml
- Scrapy





# Beautiful Soup



The most popular choice!

- BeautifulSoup is a parsing library that is used to get data out of HTML, XML, and other markup languages.
- Sometimes it is not possible to directly download the data from the websites.
- In these cases, BeautifulSoup helps you to fetch the data or the particular content by removing the HTML tags.

BeautifulSoup

# Data Analysis of Amazon Reviews using Python





# Data Analysis Online

# Data Analytics - Online



Do **not** be overwhelmed by content!

- It is key that you as a learner is not overwhelmed by the wealth of data present online.
- Working on a structured learning path is key.
- Talking with experts in the industry and discussing about learning.

# Data Analytics - Online



## Popularity of Data Analytics in college

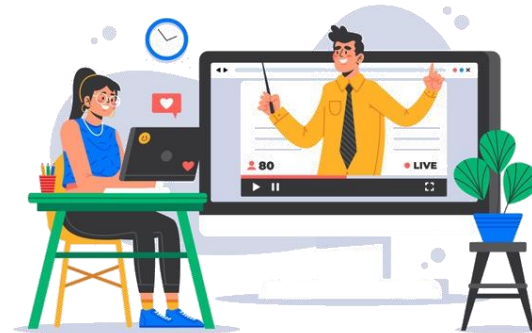
- Did you know that over 72% of college graduates say that they want to study Data Analytics online?
- Specializations help in closing on your dream designation at the coveted company.
- College students are picking up electives that help in their Data Science journey.
- Students to professionals gap is bridged by immense interest in the field.

# Data Analytics - Online



## Free Content on Data Analytics

- Great Learning **Academy**
- Great Learning **YouTube** Channel
- Great Learning **Blogs**



# Data Analytics - Online



## Specializations and Post Graduation

- PG Program in Data Science and Business Analytics – In collaboration with Texas McCombs
- PG Program in Data Science and Engineering – Great Lakes
- M.Tech in Data Science and Machine Learning – A Program by PES University
- MS In Data Science – A Program by Northwestern University

# Thank You!