upGrad

#LifeKoKaroLift

Data Science Program

Module Name: EDA & Statistics

Course : EDA

Lecture On : EDA - Day - 1

Instructor :

upGrad


E xploratory
D ata
A nalysis
21.8
shutterstock
www.shutterstock.com · 352982963

upGrad

Today's Agenda

- Introduction to EDA
- Public and Private Data
- Web Scraping
- Data Types

Data is the key

- To solve a business problem using analytics, you need to have historical data.The better the data, the more insights you can get out of it.
- When working on a new dataset in order to take intelligent action, you need to understand your data.
- Exploratory data analysis (EDA) allows us to develop the gist of what our data may look like and what kinds of questions can be answered by them.
- EDA is important because it allows the explorer to make critical decisions about what is interesting to pursue and what probably isn't worth following up on and thus building a hypothesis using the relationships between variables.

Introduction

upGrad

- Exploratory Data Analysis (EDA) is an approach/philosophy for data analysis that employs a variety of techniques (mostly graphical) to
  - Maximize insight into a data set
  - Uncover underlying structure
  - Extract important variables
  - Detect outliers and anomalies
  - Test underlying assumptions, etc.

Data Sourcing

upGrad

- Typically, data comes from various sources and your first job as a data analyst is to procure the data from them.

- The two major sources are:

    - Private Data
    - Public Data

Data Sourcing

upGrad

- A large amount of data collected by the government or other public agencies is made public for the purposes of research.
- Such data sets do not require special permission for access and are therefore called public data.
- On the other hand, private data is that which is sensitive to organisations and is thus not available in the public domain.
- Banking, telecom, retail, and media are some of the key private sectors that rely heavily on data to make decisions.

Private Data

upGrad

- A large number of organisations seek to leverage data analytics to make crucial decisions.

- As organisations become customer-centric, they utilise insights from data to enhance customer experience, while also optimising their daily processes.

- While banks use data to make credit related decisions, telecoms use data to optimise plans for customers and predict customer churn. HR data analytics helps identify and predict employee behaviour.

- While retail data analytics helps drive decisions such as pricing and stocking, the media industry uses data extensively to target viewers better. Advertisers use data to identify the best avenues for targeting customers, while journalists use data visualisation to aid information.

Public Data

upGrad

- Public data is available on the internet on various platforms.
- A lot of data sets are available for direct analysis, whereas some of the data have to be manually extracted and converted into a format that is fit for analysis.
- We have various websites which provides public (open source) data like kaggle.com, UCI Machine Learning Repository, Google Dataset Search, data.gov.in etc.
- If you are interested in sports, Awesome Public Datasets on GitHub contains a directory of sports data from tennis, cricket, football, basketball and other sports.

Key Takeaway

- Introduction to EDA
- Public and Private Data
- Web Scraping
- Data Types

upGrad

#LifeKoKaroLift

Thank You!