Group of
Horribly
Optimistic
STatisticians

Data Manipulation pt.2

Intro to Data Science

Maksymilian Norkiewicz & Jędrzej Ogrodowski

Exploratory Data Analysis

Exploratory Data Analysis (EDA) is an approach of analyzing data sets to summarize their main characteristics.

Exploratory Data Analysis

Descriptive Analysis

Exploratory Data Analysis

Descriptive Analysis

- Organizing and summarizing data using numbers & graphs
- Data Summary: Bar Graphs, Histograms, Pie Charts, Shape of graph & skewness
- Measures of Central Tendency: Mean, Median & Mode
- Measures of Variability: Range, Variance & Standard deviation

Data Mining

Data mining is the process of extracting and discovering patterns in large data sets involving methods at the intersection of machine learning, statistics, and database systems.

Data Mining

Inferential Analysis

Data Mining

Inferential Analysis

- Using sample data to make an inference or draw a conslusion of the population
- Uses probability to determine how confident we can be that the cnoclusions we make are correct. (Confidence Intervals & Margin of Error)

Storytelling

The benefits of data storytelling

- Adding value to your data and insights.
- Interpreting complex information and highlighting essential key points for the audience.
- Providing a human touch to your data.
- Offering value to your audience and industry.
- Building credibility as an industry and topic thought leader.

Making sure your data story is valuable

- Think about your theory. What do you want to prove or disprove? What do you think the data will tell you?
- Collect data. Collate the data you'll need to develop your story.
- Define the purpose of your story. Using the data you gathered, you should be able to write what the goal of your story is in a single sentence.
- Think about what you want to say. Outline everything from the intro to the conclusion.
- Ask questions. Were you right or wrong in your hypothesis? How do these answers shape the narrative of your data story?
- Create a goal for your audience. What actions would you like them to take after reading your story?

Using data visualization to enhance your data storytelling

- Reveal patterns, trends, and findings from an unbiased viewpoint.
- Provide context, interpret results, and articulate insights.
- Streamline data so your audience can process information.
- Improve audience engagement.

The three key elements of data storytelling

Build your narrative

Use visuals to enlighten

Show data to support

Data sources

- Kaggle
- Public Institutions (GUS, NYC Open Data, data.gov, etc.)
- Wikidata
- Web Scraping
- Google Dataset Search

References

- https://towardsdatascience.com/an-extensive-guide-to-exploratory-data-analysis-ddd99a03199e
- http://guidetodatamining.com/
- https://www.researchgate.net/profile/Dr-Subhendu-Pani/publication/337146539_IJITEE/links/5dc70b124585151435fb427e/IJITEE.pdf
- https://www.microsoft.com/en-sg/power-platform/products/power-bi/topics/data-storytelling
- https://www.redalyc.org/pdf/2990/299023509014.pdf
- https://en.wikipedia.org/wiki/Exploratory_data_analysis
- https://en.wikipedia.org/wiki/Data_mining
- https://www.youtube.com/watch?app=desktop&v=VHYOuWu9jQI&ab_channel=TheOrganicChemistryTutor