



DATA SCIENCE TASK OVERVIEW

MASTERING VISUALIZATION WITH PYTHON

VISUALIZATION IS NOT JUST ABOUT MAKING PRETTY GRAPHICS; IT'S A FUNDAMENTAL ASPECT OF DATA SCIENCE THAT ALLOWS US TO UNDERSTAND COMPLEX DATA SETS AND COMMUNICATE FINDINGS EFFECTIVELY. A WELL-CRAFTED VISUALIZATION CAN REVEAL INSIGHTS THAT TABLES OF NUMBERS CANNOT, MAKING IT A POWERFUL TOOL FOR ANALYSIS AND DECISION-MAKING.

Why This Task Matters:tion with Python

BECOMING PROFICIENT IN DATA VISUALIZATION TOOLS AND TECHNIQUES IS CRUCIAL FOR ANY DATA SCIENTIST. THIS TASK WILL DEEPEN YOUR UNDERSTANDING OF PYTHON'S VISUALIZATION LANDSCAPE AND IMPROVE YOUR ABILITY TO PRESENT DATA IN A CLEAR, INFORMATIVE MANNER.

BECOMING PROFICIENT IN DATA VISUALIZATION TOOLS AND TECHNIQUES IS CRUCIAL FOR ANY DATA SCIENTIST. THIS TASK WILL DEEPEN YOUR UNDERSTANDING OF PYTHON'S VISUALIZATION LANDSCAPE AND IMPROVE YOUR ABILITY TO PRESENT DATA IN A CLEAR, INFORMATIVE MANNER.

TASK REQUIREMENTS

Library Overview:

START WITH A BRIEF INTRODUCTION TO THE LIBRARIES YOU'VE CHOSEN. HIGHLIGHT WHAT SETS THEM APART AND THEIR MAIN APPLICATIONS.

Graph Types:

DETAIL THE ARRAY OF GRAPHS EACH LIBRARY CAN PRODUCE: LINE PLOTS, SCATTER PLOTS, BAR CHARTS, HISTOGRAMS, PIE CHARTS, ETC.

INCLUDE FOR EACH GRAPH TYPE: A DESCRIPTION, ITS USE CASE, AND A CODE SNIPPET SHOWING HOW TO CREATE IT.

DO DOCUMENTATION FOR ANY TWO OF MATPLOTLIB, SEABORN, PLOTLY, BOKEH, AND PANDAS

Comparison:

CONCLUDE WITH A COMPARISON OF THE LIBRARIES, FOCUSING ON EASE OF USE, CUSTOMIZATION, INTERACTIVITY, AND PERFORMANCE WITH LARGE DATA SETS. THIS WILL PROVIDE VALUABLE INSIGHTS INTO WHEN AND WHY TO USE EACH LIBRARY.

Deliverables:

- A WELL-ORGANIZED DOCUMENTATION GUIDE THAT INCLUDES THE SECTIONS OUTLINED ABOVE.
- PRACTICAL EXAMPLES WITH CODE SNIPPETS FOR EACH TYPE OF GRAPH MENTIONED.
- A THOUGHTFUL COMPARISON THAT WILL HELP OTHERS CHOOSE THE RIGHT TOOL FOR THEIR DATA VISUALIZATION NEEDS.

THIS TASK IS NOT JUST ABOUT LEARNING THE SYNTAX BUT ALSO ABOUT UNDERSTANDING THE STRATEGIC USE OF VISUALIZATION IN DATA SCIENCE. BY DOCUMENTING THESE LIBRARIES, YOU'LL NOT ONLY AID YOUR OWN LEARNING BUT ALSO CREATE A RESOURCE THAT CAN HELP THE BROADER COMMUNITY.

ASSESSING AIR PURITY: A STEP-BY-STEP GUIDE TO AIR QUALITY INDEX ANALYSIS

Air Quality Index Analysis seeks to assign a numerical figure that reflects the general state of air quality, which is crucial for the protection of public health and the stewardship of the environment. Here are the steps to undertake for conducting an Air Quality Index Analysis:

1. Compile air quality measurements from a variety of sources, including official monitoring stations, sensor data, or satellite images.
2. Perform data cleaning and pre-processing on the gathered information.
3. Utilize established equations and criteria from environmental authorities to determine the Air Quality Index.
4. Develop graphical displays, such as trend lines or heat maps, to illustrate changes in AQI over time or differences across different areas.
5. Assess the AQI levels of the area against the standard air quality benchmarks.

X SENTIMENTAL ANALYSIS

X is among the social media platforms where individuals freely express their views on various subjects. Occasionally, we witness intense debates on X regarding someone's stance, leading to an accumulation of negative posts. Bearing this in mind, if you're interested in mastering sentiment analysis on X.,

Sentiment analysis falls under the umbrella of natural language processing. All social media networks need to track the sentiments of participants in discussions. Negative sentiments are frequently observed on X, especially in political debates. Hence, each platform should persistently evaluate sentiments to identify individuals who propagate hate and negativity on their network.

For conducting sentiment analysis on X, I've sourced a dataset from Kaggle that comprises posts from an extended conversation among a group of users. Our objective here is to determine the number of negative versus positive posts, enabling us to conclude.

- **VADER Sentiment Analyzer:**

- Interns are encouraged to use the VADER (Valence Aware Dictionary and sEntiment Reasoner) Sentiment Analyzer.
- A lexicon and rule-based sentiment analysis tool that is specifically attuned to sentiments expressed in social media.
- Offers simplicity and high precision in analyzing sentiments of texts.

- **Task for Interns:**

- Use the VADER Sentiment Analyzer to evaluate sentiments of posts on X.
- Aim to quantify negative and positive sentiments from the provided dataset.
- Draw conclusions based on the sentiment analysis to understand the overall sentiment trend.

TASK 3

CRICKET FIELDING ANALYSIS DATA COLLECTION

- **Objective:** Conduct detailed fielding performance analysis for three players from any innings of a T20 match.
- **Data Collection:** Record fielding efforts for each ball bowled, categorizing actions and outcomes.
- **Analysis Preparation:** Organize collected data for advanced fielding analysis, identifying strengths and areas for improvement.
- **Deliverable:** Compile fielding data into a spreadsheet or database, contributing to strategic fielding placements and team performance enhancements.

JUPYTER NOTEBOOK DATA ANALYSIS PROJECT

- Objective: Create a Jupyter notebook from scratch and conduct data analysis on a dataset of choice.
- Dataset Selection: Choose an intriguing dataset and explore its contents within the notebook.
- Research Question Formulation: Define a focused research question based on dataset exploration.
- Visualization and Insights: Utilize various visualization techniques to derive meaningful insights and solutions to the research problem.

EXAMPLE SCENARIO: SALES DATA ANALYTICS

- Dataset Exploration: Dive into transactional intricacies including gross sales, net sales, profit/loss, COGS, manufacturing costs, and freight costs.
- Research Question Formulation: Develop nuanced questions exploring correlations between financial variables, such as manufacturing costs and net sales.
- Visualization Techniques: Implement sophisticated financial visualizations, such as cost breakdowns and profit margin trends, using Python libraries like Matplotlib and Seaborn.
- Jupyter Notebook Presentation: Present temporal trends in sales metrics alongside detailed breakdowns of financial variables, showcasing proficiency in data analytics and financial reporting nuances.

THANK YOU!

