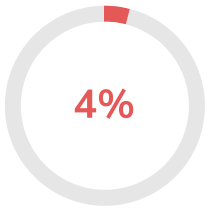
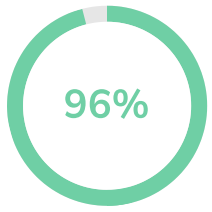
Plagiarism Scan Report

Report Generated on: Mar 27,2024



Plagiarised



Unique

Total Words:	**911**
Total Characters:	**6725**
Plagiarized Sentences:	**2.72**
Unique Sentences:	**65.28 (96%)**

Content Checked for Plagiarism

Abstract

This project delves into the analysis of an e-commerce dataset using the Pandas library in Python, conducted within a Jupyter notebook environment. The dataset includes information such as purchase prices, credit card providers, job titles, and languages of customers. The analysis begins with basic exploratory tasks, such as displaying the top and last rows, checking for null values, and summarizing the dataset's structure.

Further analysis includes identifying trends such as the highest and lowest purchase prices, the average purchase price, and the distribution of purchases by language and job title. Intermediate-level analysis involves filtering data to find specific patterns, such as customers with Mastercard as their credit card provider who made purchases above a certain threshold or those with credit cards expiring in a particular year.

Data visualization plays a crucial role in this analysis, with various plots illustrating purchase counts, purchase prices, and company purchases. Scatter plots highlight purchase prices based on specific job titles, while bar charts showcase average purchase prices by language and browser usage. These visualizations offer insights into customer behavior and preferences within the e-commerce platform.

This project demonstrates the versatility of Pandas in analyzing and visualizing complex datasets, providing valuable insights for e-commerce businesses to understand their customer base better and make informed decisions.

Introduction

The proliferation of e-commerce platforms has led to an abundance of data that provides gives insights .
In this project, we analyze the 'E-Commerce Purchase' dataset, downloaded from Kaggle, using the powerful data manipulation and analysis library, Pandas, in Python. This dataset contains information on e-commerce purchases, including details such as purchase prices, credit card providers, job titles, and languages of customers etc.

Our objective is to conduct a comprehensive analysis of this dataset, exploring various aspects such as purchase trends, customer demographics, and popular products. By leveraging the capabilities of Pandas, we aim to extract meaningful insights that can help e-commerce businesses make

informed decisions and enhance their understanding of customer preferences.

Through this analysis, we aim to demonstrate the effectiveness of Pandas in handling and analyzing real-world datasets, highlighting its utility in extracting valuable insights from complex data. The findings of this analysis have the potential to inform marketing strategies, product offerings, and customer engagement initiatives, ultimately contributing to the growth and success of e-commerce businesses.

What is Pandas?

Pandas is an open-source data that manipulation and analysis library for Python. It gives us easy-to-use data structures and functions designed to make working with structured data fast, easy, and expressive. Pandas is built on top of NumPy, another Python library that provides support for large, multi-dimensional arrays and matrices, making it a powerful tool for data analysis and manipulation.

The primary data structures in Pandas are Series and DataFrame. A Series is a one-dimensional array-like object that can hold various data types, such as integers, strings, or floating-point numbers. A DataFrame is a two-dimensional, size-mutable, and heterogeneous tabular data structure with labeled axes (rows and columns).

Pandas provides a wide range of functions for reading and writing data, data cleaning, reshaping, merging, slicing, indexing, and aggregating data. It also integrates seamlessly with other libraries in the Python ecosystem, such as Matplotlib for data visualization and scikit-learn for machine learning tasks.

In this project, we leverage the power of Pandas to explore and analyze the 'E-Commerce Purchase' dataset, demonstrating its capabilities in handling and analyzing real-world data efficiently.

• Importing Libraries

• Reading the whole Dataset

Basic things

• Reading first 10 column in the Dataset

• Reading the last 10 columns in the Dataset

Data Cleaning

Data cleaning is a crucial step in the data analysis process, as it ensures that the dataset is accurate, consistent, and ready for analysis.

• Handling missing values

Here we verified whether any null values (missing values) present the "E-commerce Purchase" Dataset.

Exploratory Data Analysis

Exploratory Data Analysis is an crucial step in the data analysis process, as it helps us understand the structure of the dataset, identify patterns, and generate hypotheses for further analysis. In this project, we conducted a thorough EDA of the 'E-Commerce Purchase' dataset, using Pandas to perform basic and intermediate analyses.

I. Basic Analyses:

▪ We explored the number of columns and rows in the dataset using the len() function.
▪ We identified the columns present in the dataset using the columns attribute.

▪ We calculated summary statistics such as the highest, lowest purchase prices.

▪ We counted the number of people with French as their language and job titles containing "Engineer."

II. Intermediate Analyses:

▪ We identified people with Mastercard as their credit card provider who made purchases above a certain threshold.

▪ We counted the number of people with a credit card that expires in 2020.

Data Visualization

Data visualization plays an important role in understanding difficult datasets, as it allows us to visually explore patterns, trends, and relationships . In this project, we used various data visualization techniques to enhance our understanding of the 'E-Commerce Purchase' dataset and communicate our findings effectively.

Key Findings

After conducting a comprehensive analysis of the 'E-Commerce Purchase' dataset, several key findings emerged, shedding light on customer behavior and trends within the e-commerce platform.

Finally ,these key findings showed us valuable insights into customer preference and preferences within the e-commerce platform, offering opportunities for customer engagement initiatives.

Conclusion

▪ In conclusion, the analysis of the 'E-Commerce Purchase' dataset has showed us the potential of data analysis and visualization in extracting meaningful insights from complex datasets.

23+ Pandas Interview Questions: Dive into Data Science

Apr 23, 2023 — The primary data structures in Pandas are Series and DataFrame. A Series is a one-dimensional labeled array capable of holding any data type ...

https://seterx.com/pandas-interview-questions

56%

How To Create A Dataframe In Python: Step-By-Step Instructions

Oct 4, 2023 · A dataframe is a two- dimensional, size-mutable, and heterogeneous tabular data structure with labeled axes (rows and columns) in pandas. It's similar to a spreadsheet or SQL table and is one of the most commonly used pandas objects.

https://marketsplash.com/tutorials/python/how-to-create-a-dataframe-in-python

44%

What is Data Visualization and Its Importance?

Oct 12, 2023 — Data visualization plays an important role in business analytics because it helps businesses see the latest trends and patterns. For example, if ...

https://www.knowledgehut.com/blog/business-intelligence-and-visualization/importance-of-data-visualization

49%