UDACITY Google spark

Palestine Launchpad Program

Data Analysis NanoDegree

Our agenda for today

1. Recap about the last session
2. QA: Pareto Diagram (Explained)
3. Project (Three) – Part 1 Exploratory Data Analysis
4. Case Study
5. New Concepts (Fundamentals of accelerating Data)
6. Knowledge Sharing (Text-Processing)
7. Q\A

Palestine Launchpad
منصة فلسطين التقنية
Google

Recap

What did we talk about last session

UDACITY

Last Session we covered the following:

✓ Project 3 Introduction and Discussion

✓ 🚀 We introduced and discussed the details of your third project.
🔍 Part 1 focuses on the dataset exploration phase.
📊 We agreed on choosing the dataset and starting the exploration phase.

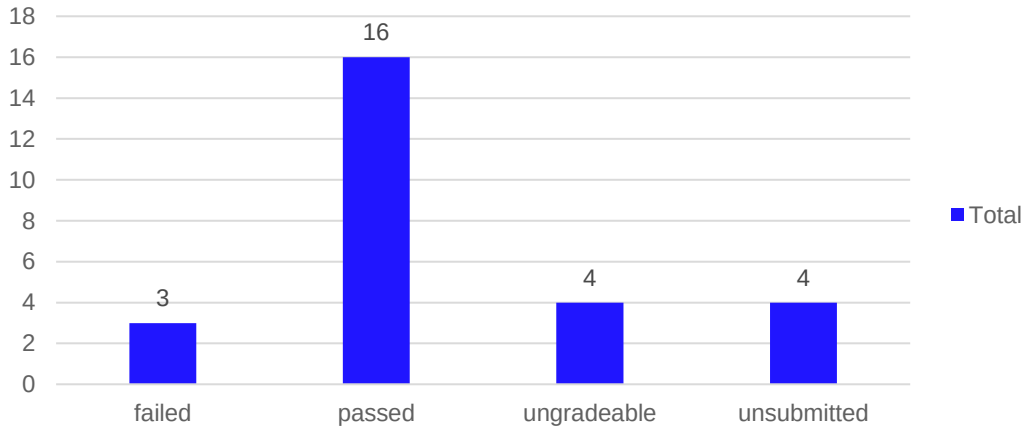
Data Classification Methods:

✓ 🧠 Explored different data classification methods, emphasizing the importance of data types and measurement levels.
📈 Highlighted the best charts for categorical data types and illustrated these concepts using a Colab notebook.

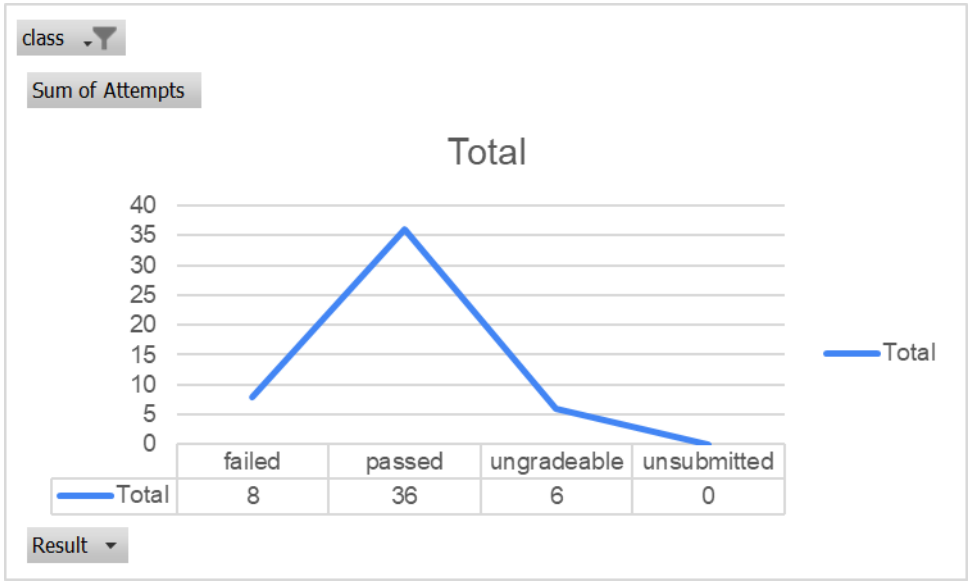
Fundamentals of Accelerated Data Science:

✓ ⚡ Talked about key concepts and techniques in accelerated data science with a running example using RAPIDS and cuDF

Project Two Statistics Update



Project Two Statistics Update



Palestine Launchpad
منصة فلسطين التقنية
Google

QA: Pareto Explained

What did we talk about last session

UDACITY

Visualizing Data (Categorial Data Types) :

1. Frequency distribution table (Table with frequencies)
2. Bar Charts
3. Pie Charts (Percentage (Share of the total))
4. Pareto Diagram (Special Bar Chart)

https://colab.research.google.com/drive/1rNcwrSXRCnnEuRfVjzzd5HrL1L1izboc?usp=sharing

Pareto Diagram: Demo

https://colab.research.google.com/drive/1FQtkWxjGAPVhymXbRVTbcjFWw7Ibt-AT?usp=sharing

Palestine Launchpad
منصة فلسطين التقنية
Google

Project Three : Communicate Data Findings

What is needed to make a successful submit

UDACITY

Project: Communicate Data Findings

Project Overview
This project has two parts that demonstrate the importance and value of data visualization techniques in the data analysis process.

- ✓ **In Part I, Exploratory data visualization**, you will use Python visualization libraries to systematically explore a selected dataset, starting from plots of single variables and building up to plots of multiple variables.

- ✓ **In Part II, Explanatory data visualization**, you will produce a short presentation that illustrates interesting properties, trends, and relationships that you discovered in your selected dataset. The primary method of conveying your findings will be through transforming your exploratory visualizations from the first part into polished, explanatory visualizations.

- ✓ **Project Due Date: Aug 27, 2024**

Choose Your Dataset

Below is a compiled list of the datasets you can choose from. However, you can explore other datasets that interest you:

- ❑ **Ford GoBike System Data(opens in a new tab)** (38 MB, CSV File)
- ❑ **Flights**
- ❑ **Loan Data from Prosper(opens in a new tab)** (82.5 MB, CSV File)
- ❑ **PISA Data - 2018**

Pareto Charts & 80-20 Rule - Clinical Excellence Commission (nsw.gov.au)

Univariate Exploration

- Histogram - required
- Bar Chart
- Count Plot

Bivariate Exploration

- Scatterplots - required
- Box Plots - required
- Clustered Bar Chart
- Heatmap

Multivariate Exploration

- Facet Plot - required
- Plot Matrix
- Scatterplot with multiple encodings

Project Three Rubric –PART ONE

Is the data explored systematically using a series of appropriate and varied visualizations?

Student has provided seven (7) exploratory data visualizations distributed over univariate, bivariate, and multivariate plots to explore many relationships in the data set.

For each exploration **category**, complete the required chart(s) and choose one additional visualization type.

Project Three Rubric –PART ONE

Are questions and observations documented in the report?

Questions and observations are placed regularly throughout the report.

- For each exploration category (univariate, bivariate, multivariate) state the assumptions and questions the charts should answer.
- After each plot or set of related plots, describe what you found.

Tip: Use the ""Question-Visualization-Observations"" framework throughout the exploration.

Tip: For the Part I notebook, use *File* > *Download as...* > *HTML or PDF* menu option to generate the HTML/PDF.

Histogram in Seaborn

- https://colab.research.google.com/drive/1nJy6Us1zOOH6Qr46f2enhBCV-deALUG0?usp=sharing

Palestine Launchpad
منصة فلسطين التقنية
Google

New Concepts

Udacity, Students and Session Leads

UDACITY

SELENIUM WEBDRIVER

https://selenium-python.readthedocs.io/

1. Install selenium library
2. Install WebDriver suitable for your browser

Palestine Launchpad
منصة فلسطين التقنية
Google

Case Study

Viulize a log database

UDACITY

EXPLORE PROJECT 3 DATASETS

Palestine Launchpad
منصة فلسطين التقنية
Google

Knowledge Sahring

Udacity, Students and Session Leads

UDACITY

Text processing pipeline by Sondos

Palestine Launchpad
منصة فلسطين التقنية
Google

Next Steps

Let's get things started...

UDACITY

Second Project July 2, 2024

May | June | July | August | September

Strat working on Project three

1. Choose the dataset
2. Start Exploring your dataset

Today
July 6, 2024

Third Project Due Date

August 27, 2024
- Communicate Data findings

Program Period
May 9, 2024 - September 14, 2024

Palestine Launchpad
منصة فلسطين التقنية
Google

Thank You!

And Good Luck!

UDACITY