K Youth Data Science Bootcamp

Capstone Project

In this project, you will be given the opportunity to show off the skills that you have learned. You will need to make a 5 – 7 minute presentation on a topic of your choice showing how you implement the data science workflow to create a data science model.

You may select any suitable data set that reflects your field of study, interest or area of work to address a data science problem in this area. Do include background and explanations within the Jupyter notebook for your presentation and submission.

Your work should be compiled in a Jupyter notebook with the following content and related code.

- Introduction
    - Problem Statement
    - Objectives
- Data Gathering and Preparation
    - Sources of Data
    - Data wrangling steps
- Exploratory Data Analysis (EDA)
    - Basic descriptive statistics and visualizations
- Modeling
    - Explain the modeling technique to solve the problem, including choice of model and parameters
    - Explain the insights and results obtained
- Evaluation and Recommendations
    - Evaluation of the model using appropriate metrics
    - Provide the recommendations according to the initial problem statement.

You will be required to submit the notebook and the related datasets, into the Google Classroom for the course.

K Youth Data Science Bootcamp

Project Rubrics

Your instructor will review your project and evaluate it against the rubric below. You will receive a score of 0-25 for each criterion:

- 0: Missing/incomplete
- 15: Does not meet expectations
- 20: Meets expectations
- 25: Exceeds expectations

Each category will have a maximum of 25 points. Your grade will be the sum of these scores, with 100 points being the highest possible result. You need to score at least 80 out of 100 to earn a passing grade.

Problem Statement and Definition • Develops well-informed, hypothesis-driven business questions for analysis. • The problem statement should be specific and to the point, and it should not be vague or ambiguous • Defines the problem. The problem statement should define the problem in detail. This means explaining the causes of the problem, the effects of the problem, and the scope of the problem.	/ 25
Data Gathering and Preparation • Identify the data needs. The first step is to identify the data that is needed for the project. This will depend on the specific goals of the project. • Gather the data. Once the data needs have been identified, the data can be gathered from a variety of sources, such as databases, surveys, and social media. • Clean the data. The data that is gathered need to be cleaned to remove errors and inconsistencies. • Prepare the data for analysis. Data need be prepared for analysis by transforming it into a format that can be used by the chosen analytical tools. This may involve standardizing the data, creating new features, or removing outliers. •	/ 25
EDA and Modeling • Provides relevant descriptive statistics on the data • Creates appropriate visualizations to explore the data. • Data Modeling: Choose the right modeling technique to solve the problem.	/ 25
Evaluation and Recommendations • Identifies key insights pertaining to the original business question(s). • Incorporates results of the analysis to support decisions and recommendations.	/ 25