/ [Professional Certificate in Machine Learning and Artificial Intelligence - October 2024](https://classroom.emeritus.org/courses/9464) / [Modules](https://classroom.emeritus.org/courses/9464/modules) / [Required Capstone Assignment 20.1: Initial Report and Exploratory Data Analysis (EDA) [7:00:00]](https://classroom.emeritus.org/courses/9464/assignments/293119)

**Required Capstone Assignment 20.1: Initial Report and Exploratory Data Analysis (EDA) [7:00:00]**

* Due Apr 16 by 4:29pm
* Points 25
* Submitting a website url
* Attempts 0
* Allowed Attempts 1
* Available after Apr 9 at 1pm

*There is one required assignment on this page. Please scroll down to complete it.*

**Learning Outcome Addressed**

1. Perform a deep analysis of your identified problem in order to develop an executive brief detailing your technical findings.

***This is a required assignment and counts toward program completion.***

**Overview**

In this module, you will work on performing exploratory data analysis (EDA) to develop an initial report for your capstone project. You will use EDA to see what data can reveal beyond the formal modeling, hypothesis testing task, and data training to provide a better understanding of dataset variables and the relationships between them. You are encouraged to spend your time in this module cleaning your data and use feature engineering and EDA techniques to create visualizations to make sense of your findings. Additionally, you will also be required to use one of the ML algorithms you have learned so far in the program to develop a baseline model to use as a comparison in Module 24. You will have time in Module 24 to include additional models, clean the code, and make your work presentable for technical and non-technical audiences. For now, you will do the ‘heavy lifting’ of finding the answer to your research question.

The only deliverables for this module will be the Jupyter notebook(s) where you did your analysis and the updated [README.md](https://classroom.emeritus.org/courses/9464/files/4026708?wrap=1)

[Download README.md](https://classroom.emeritus.org/courses/9464/files/4026708/download?download_frd=1) file with the results section completed. If you have changed methodologies or datasets since Module 16, please make sure you reflect those changes in the [README.md](https://classroom.emeritus.org/courses/9464/files/4026706?wrap=1)

[Download README.md](https://classroom.emeritus.org/courses/9464/files/4026706/download?download_frd=1) file as well.

**Submission Instructions:**

* Post both your [README.md](https://classroom.emeritus.org/courses/9464/files/4026707?wrap=1)
* [Download README.md](https://classroom.emeritus.org/courses/9464/files/4026707/download?download_frd=1) file and your Jupyter notebooks to your GitHub Repository. Submit a link to your repository here.

*The average completion time for this assignment is 15 hours.*

**Grading:**

Please note that you will be awarded points based on this activity. You may refer to the rubric below to know more about the grading criteria.

* [Website URL](https://classroom.emeritus.org/courses/9464/assignments/293119?module_item_id=2259974#submit_online_url_form)

Top of Form

|  |
| --- |
| Copy and paste the link to the web site you'd like to submit for this assignment. |
| Website URL: |
|  |
|  |

Bottom of Form

**Rubric**

Required Capstone Assignment 20.1: Initial Report and Exploratory Data Analysis (EDA)

| Required Capstone Assignment 20.1: Initial Report and Exploratory Data Analysis (EDA) | | |
| --- | --- | --- |
| **Criteria** | **Ratings** | **Pts** |
| This criterion is linked to a Learning Outcome Project organization:  • Word document with a summary of findings and a link to the Jupyter notebook. • Jupyter notebook has headings and text appropriately formatted. • No unnecessary files. • Directories and files have appropriate names and locations. | |  |  | | --- | --- | | 5 pts  Excellent  Your submission includes all of the listed components. | 0 pts  Criterion Not Met  Your submission includes a few or none of the listed components. | | 5 pts |
| This criterion is linked to a Learning Outcome Syntax and code quality:  • Libraries are imported and aliased correctly. • Code does not contain errors. • No long strings of code output. • Demonstrates competency with pandas. • Demonstrates competency with visualization libraries e.g., seaborn, Matplotlib, and Plotly. • Comments are used appropriately to explain code. • Variables are sensible. | |  |  | | --- | --- | | 5 pts  Excellent  Your submission includes all of the listed components. | 0 pts  Criterion Not Met  Your submission includes a few or none of the listed components. | | 5 pts |
| This criterion is linked to a Learning Outcome Visualizations:  • Appropriate plots for categorical and continuous variables are utilized. • Plots contain readable labels. • Plots contain descriptive titles. • Axes are legible. • Subplots are used when appropriate. • Plots are scaled appropriately for readability. | |  |  | | --- | --- | | 5 pts  Excellent  Your submission includes all of the listed components. | 0 pts  Criterion Not Met  Your submission includes a few or none of the listed components. | | 5 pts |
| This criterion is linked to a Learning Outcome Data Cleaning and EDA  • The dataset is clean: imputation and removal of missing values, removal of duplicate values • Outliers analysis to identify anomalies in the dataset. • Perform feature engineering to extraction and transformation of variables from raw data. | |  |  | | --- | --- | | 5 pts  Excellent  Your submission includes all of the listed components. | 0 pts  Criterion Not Met  Your submission includes a few or none of the listed components. | | 5 pts |
| This criterion is linked to a Learning Outcome Modeling:  • Identify an appropriate classification or regression ML model to utilize as baseline for your analysis. • Valid interpretation of an evaluation metric. • Clear identification of an evaluation metric. • Clear rationale for use of the given evaluation metric. | |  |  | | --- | --- | | 5 pts  Excellent  Your submission includes all of the listed components. | 0 pts  Criterion Not Met  Your submission includes a few or none of the listed components. | | 5 pts |
| Total Points: 25 | | | |

The following content is partner provided

The preceding content is partner provided