ASHESI

Masters in Intelligent Computing Systems
Machine Learning and Data Science (September 2025)

Problem Situation 1 (PROSIT 1): Charting the Path to Student Success

Ashesi University has long prided itself on graduating ethical, entrepreneurial leaders prepared to transform Africa. Yet, as the student body grows more diverse, the leadership at Ashesi have begun asking a difficult question:

What factors truly influence whether students thrive at Ashesi?

Each year, Ashesi admits students from a wide range of high schools, regions, and backgrounds. Some adjust quickly and excel academically, while others struggle with coursework, retention, or progression. The leadership at Ashesi suspects that there are patterns in the data that could help them better understand these outcomes, but the patterns are not obvious.

To explore this, they have committed to sharing a dataset that combines admissions records with registry data by Wednesday 10th September 2025 . They have asked you and your team to investigate the dataset and provide an initial report. But they've made one request: "Please don't jump straight into building models. First, help us see and understand the data clearly."

Your Deliverables

- Script used for data exploration and pre-processing
- In-class presentation of process and findings

Learning Outcomes

- Articulate the role of exploratory data analysis (EDA) in the Machine Learning pipeline
- Distinguish among data types and choose appropriate summary statistics and visualisations for each.
- Compute and interpret key numeric descriptors such as variance, standard deviation, inter-quartile range.
- Create and critique graphical summaries such as box plots and heatmaps

- Describe common probability distributions (Normal, Binomial, Poisson, Exponential) and explain sampling distributions and the Central Limit Theorem (CLT) in the context of estimating population parameters from samples.
- Perform core data-quality and preprocessing steps such as detecting/handling missing values and outliers, type casting, encoding categoricals, scaling and normalisation, and justify each choice.
- Recognize and articulate the ethical considerations in working with sensitive data, including issues of privacy, confidentiality, fairness, and potential bias.
- Propose strategies to ensure that exploratory analyses and preprocessing steps do not reinforce stereotypes or harm underrepresented groups.

Resources

Textbooks

Anand, G., & Sharma, R. (2022). *Data science fundamentals and practical approaches.* BPB Publications.

Gupta, P. (2022). *Practical data science with Jupyter.* BPB Publications.

Ozdemir, S. (2024). *Principles of data science: A beginner's guide to essential math and coding skills for data fluency and machine learning* (3rd ed.). Packt Publishing Ltd.

Jupyter Notebooks

Healthcare Dataset - EDA: https://www.kaggle.com/code/hainescity/healthcare-dataset-eda

No-Show Appointment Exploratory Data Analysis:
https://github.com/jemc36/Udacity-DAND-EDA-NoShowApp

Exploratory Data Analysis: Customer Loans in Finance:
https://github.com/joelsud18/exploratory-data-analysis---customer-loans-in-finance

Exploratory-Data-Analysis-EDA-in-Banking-Using-Python:
https://github.com/SouRitra01/Exploratory-Data-Analysis-EDA-in-Banking-Python-Project-

Articles

Exploratory Data Analysis with Python Jupyter Notebook: A tutorial on how to perform exploratory data analysis (EDA) in Jupyter Notebook, covering data cleaning, data preprocessing & data visualization techniques:
https://medium.com/%40techlatest.net/exploratory-data-analysis-with-python-jupyter-notebook-a-tutorial-on-how-to-perform-exploratory-5a800791b04f

Quick Guide to Exploratory Data Analysis Using Jupyter Notebook:
https://www.geeksforgeeks.org/data-analysis/quick-guide-to-exploratory-data-analysis-using-jupyter-notebook

A Quick Guide to Exploratory Data Analysis Using Jupyter Notebook:
https://saturncloud.io/blog/a-quick-guide-to-exploratory-data-analysis-using-jupyter-notebook