Developer Salary & Student Performance – Streamlit App

This Streamlit app is a small data storytelling portfolio project. It explores developer salaries and student dropout risk through interactive visuals, giving visitors the opportunity to ask questions about the data and see patterns at a glance.

Author & Contact

- **Name:** Ben Hislop
- **LinkedIn:** https://www.linkedin.com/in/benjaminhislop/
- **GitHub:** https://github.com/Bphissles

App Navigation Overview

The Streamlit app is organized into the following files/pages:

- `Bio.py`
  Landing page with project introduction and navigation links.
- `pages/1_EDA_Gallery.py`
  EDA Gallery for the developer salary dataset (boxplots, histogram, trend line, remote vs on-site bar chart, sidebar filters, color theme).
- `pages/2_Dashboard.py`
  Student performance dashboard using the dropout-risk dataset (scatter, heatmaps, filters, color theme).
- `pages/3_Future_Work.py`
  Future work and next-steps notes.

Datasets

1. Developer Salaries (EDA Gallery)

- **Dataset Name:** Data Developer Salary in 2024
- **Source:** https://www.kaggle.com/datasets/shahzadi786/11111111111111111111
- **Last Updated:** 2025-11-14 (as of this project)
- **Number of Rows:** 16,534

Preprocessing / Cleaning

- Loaded directly from CSV into pandas.
- Basic type coercion and filtering are handled via Streamlit filters (experience level, work year, remote ratio).

Ethics Note

- This dataset contains salary information for individual roles, not directly identifiable people; it is used only for aggregated visualization.
- All analysis focuses on distributions and high-level trends rather than individual records.
- Any conclusions are descriptive, not prescriptive, and should not be treated as career or compensation advice.

2. Predict Students' Dropout and Academic Success (Dashboard)

- **Dataset Name:** Predict Students' Dropout and Academic Success
- **Source:** UCI Machine Learning Repository – https://archive.ics.uci.edu/dataset/697/predict+students+dropout+and+academic+success
- **Last Updated:** 2021-12-12 (per source)
- **Number of Rows:** 4,424

Preprocessing / Cleaning

- Loaded from the raw CSV file using the documented semicolon (`;`) delimiter.
- Visualizations use existing numeric and categorical fields directly (admission grade, semester grades, units enrolled/approved, unemployment rate, target outcome, etc.).
- No rows were manually removed for the dashboard; where filters are applied (by outcome or age), they are done dynamically in the app.

- Data exploration here: https://github.com/Bphissles/cs3120-final-project/blob/main/research-space/final-project-milestone.ipynb confirmed no missing values.

Ethics Note

- The dataset describes student demographics and academic outcomes, which are sensitive topics.
- Visuals are used only to summarize patterns; the app intentionally avoids causal claims (e.g., “X causes dropout”) and instead focuses on what the data shows.
- No attempt is made to identify individual students. The project is for educational and exploratory purposes, not for high-stakes decision-making.

Requirements

Python dependencies are listed in `requirements.txt` at the root of this Streamlit project. The key libraries include:

- **streamlit** – app framework and layout
- **pandas** – data loading and wrangling
- **plotly** – interactive charts and hoverable visuals
- **matplotlib** (for any legacy/static charts)

To run locally:

```
pip install -r requirements.txt
streamlit run Bio.py
```

You can then navigate between pages using the links at the bottom of each page or via the Streamlit sidebar menu.

AI Assistance

This project was developed with the assistance of generative AI tools for code structure, debugging, and documentation refinement.