Does a higher academic degree guarantee higher income?

Project Description: Education and income are two of the most important factors influencing an individual's quality of life and societal status. Understanding the relationship between academic attainment and earning potential provides key factors for decision making especially for younger audiance.

Project Questions

- Does having a higher academic degree guarantee higher income?
- Can you still make alot of money even without going to university?
- Is university waste of time as proclaimed by social media gurus?

Data Sources

Data Source 1: *American Citizens Annual Income*	**Data Source 2:** *Global Salary DataSet 2022*
URL	**URL**
Data type: *CSV*	**Data type:** *CSV*
License: CDLA-Sharing-1.0	**License:** CDLA-Sharing-1.0
Source	**Source**

Data Source 1 Description:	Data Source 2 Description:
This dataset provides detailed information about income levels and various demographic attributes for individuals in the United States. It contains fields such as age, education level, country of origin, and income brackets (e.g., >50K and <=50K). The dataset is valuable for studying how personal and demographic factors influence earning potential, particularly the role of education.	This dataset offers global insights into salary trends and transparency. It includes information about education levels, country of employment, and salary ranges, allowing us to study income variations across different regions and academic qualifications. This dataset complements the American Citizen Income Dataset by providing a broader, country-level perspective.

Pipeline

Extraction	Technology used for the data extraction is Python. Both of the data sources were downloanded from Kaggle using the KaggleAPI .
Cleaning	Rows with missing or placeholder values (like ?) were filtered out using **pandas** in Python. Unnecessary columns were dropped to focus on key variables like education, country, and income.
Transformation and Integration	Income categories were standardized to binary classifications (>=50K and <50K) for uniformity across datasets. The cleaned datasets were merged based on shared attributes (e.g., education level, country) to create a unified dataset.
Storage	Stored the unified dataset in the /data directory as an SQLite file (data.sqlite)
Further analysis	The cleaned and integrated data will be analyzed to identify trends and correlations between education levels and income.

Results and Limitations

Limitations

Data Accuracy:

- The accuracy of the analysis is heavily dependent on the quality of the source data. Errors or inconsistencies in the original datasets might propagate through the pipeline.

Generalizability:

- While the datasets provide mostly insights about North Americas, results may not generalize for South America, since there is less data present from the data sources and also due to differences in economic structures and education systems.

Results

The output of the pipeline is a **SQLite database** stored in the /data folder. This database contains a fully integrated dataset combining information about academic degrees, income levels, and country-specific literacy and salary statistics.

The structure of the final dataset is clear and consistent, with missing or invalid data (e.g., fields with ?) successfully removed.

The data is now ready for analysis, with fields aligned to explore the relationship between education and income levels in the Americas.

The chosen format ensures easy querying and supports efficient analysis using tools like Python or SQL-based applications.

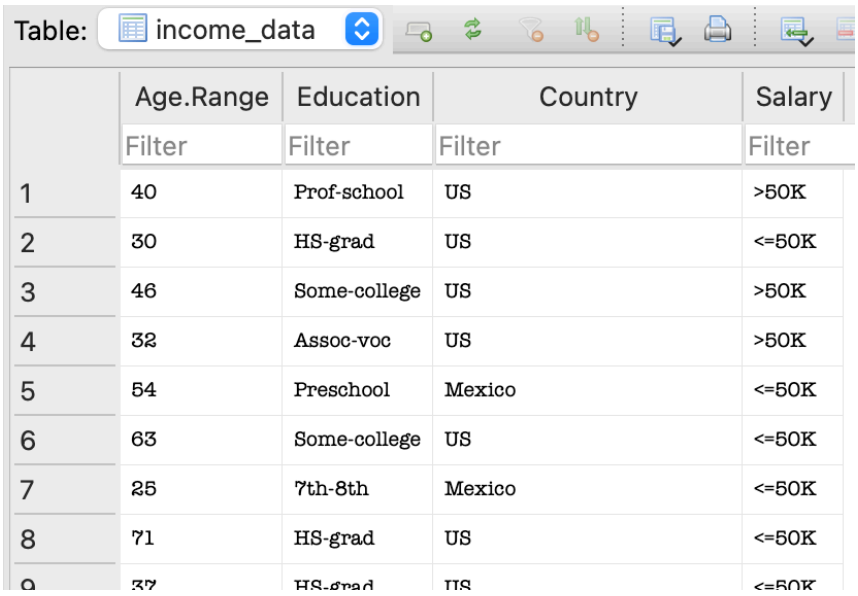

Table: income_data

	Age.Range	Education	Country	Salary
	Filter	Filter	Filter	Filter
1	40	Prof-school	US	>50K
2	30	HS-grad	US	<=50K
3	46	Some-college	US	>50K
4	32	Assoc-voc	US	>50K
5	54	Preschool	Mexico	<=50K
6	63	Some-college	US	<=50K
7	25	7th-8th	Mexico	<=50K
8	71	HS-grad	US	<=50K
9	37	HS-grad	US	<=50K

Table 1: Showing few rows of the final transformed dataframe