# Data Visualization: Histogram

## Project Overview

This project focuses on data visualization using histograms. It utilizes a SQLite database file and the Pandas library to extract and visualize data in a meaningful way.

## Features

* Downloads a SQLite database file.
* Connects to the SQLite database.
* Extracts relevant data using SQL queries.
* Uses Pandas for data manipulation.
* Generates histograms to visualize data distribution.

## Technologies Used

* Python
* Pandas
* SQLite
* Matplotlib/Seaborn (for visualization)

## Data Cleaning:

* Removal of missing values (nan) using `.dropna`
* Removal of empty rows
* Spitting of data for column `DatabaseWantToWorkWith` using `.str.split(';')` to separate the information joined by ‘;’.

## Normalization:

* `CompTotal` column was normalized using a logarithmic transformation

## Data encoding

* Encoding `YearsCodePro` into required bins for better data handling.

## Analysis & results:

### Plot 1: Coding Experience Distribution

This histogram analyzes the distribution of respondents' years of coding experience. The majority of respondents have less than 10 years of experience, with very few having over 50 years in the field.

### Plot 2: Compensation Across Age Groups

This analysis examines compensation trends across different age groups. Respondents aged 25-34 receive the highest compensation, while those aged 65 and above receive the least. This trend may be attributed to the higher number of coders in the 25-34 age bracket, leading to increased compensation opportunities.

### Plot 3: Time Spent Searching for Information

This visualization explores the time different age groups spend searching for information online. The 25-34 age group exhibits the highest engagement in online research.

### Plot 4: Preferred Databases for Future Learning

Among the surveyed respondents, PostgreSQL emerged as the most desired database for future learning and usage.

### Plot 5: Preferred Work Arrangements

Since the majority of respondents belong to the 25-34 age group, an analysis was conducted to determine their preferred work arrangement. Hybrid work models were the most favored, likely due to the flexible nature of coding, which does not require physical office presence.

### Plot 6: Mid-Career Compensation Trends

Building on Plot 2, further analysis was conducted to examine compensation trends among mid-career professionals. Individuals aged 45-54 receive higher compensation compared to those aged 55-64.

### Plot 7: Job Satisfaction Based on Experience

This analysis explores whether job satisfaction correlates with compensation and years of experience. The findings indicate that respondents with 50+ years of coding experience, although few in number, report the highest average job satisfaction. Additionally, satisfaction levels tend to increase with experience.

## Conclusion

This project provides meaningful insights into coding experience, compensation trends, and work preferences using histogram-based visualizations. The findings highlight the prevalence of younger coders in the workforce, variations in compensation across different age groups, and preferred databases and work environments. The analysis suggests that experience positively correlates with job satisfaction and that hybrid work models are widely preferred. This study demonstrates the effectiveness of histograms in data-driven decision-making and trend analysis.

## Future considerations

A time-serries study to track how coding experience, compensation, and job satisfaction evolve over time. (with appropriate data)

Compare inter-industry compensation trends and job satisfaction among developers working in different sectors (e.g., tech, finance, healthcare, gaming).