For week 35:

Dataset has 608 columns, where many of them are part of a multi-column ‘dummies’ format groups, which uses separate columns for each unique answer to the same question. This is why the dataset is so wide

Polars concat\_list was used to put all of the dummy column answers into a column of type list (polars support for list and array types if far more extensive and efficient than the pandas equivalent). From there I calculate the number of answers from each participant to selected questions and add weighting factors that decreased the value of each answer in proportion to the number of answers given. For example, when users list 4 programming languages, each one is given a weight of ¼, when users list 2 programming languages each one is given a weight of ½. An index is added to give each user a unique ID, and from there the dataframe is exploded to put each of the user choices on to separate rows. This is like a melt or unpivot.

The cleaned dataset is saved as a parquet file, with all datatypes optimized for storage size. String categories have all been cast as Categorical, float columns are all Float32, Integer columns are all UInt8. The app checks for the parquet file. If found it will read it an use for the dataset. If not found, it will read th csv file, clean it, and save the data frame as parquet for future runs. In this way I only upload a relatively small parquet file to plotly studio.

Deciding what to include in the dashboard and cleaning the dataset were formidable tasks.

This dashboard focuses on gender specific attributes for number of users by country, languages spoken at work, favorite computer language, favorite AI assistants, and favorite AI features. A global filter allows users to specify a specific country, or include them all.

TODO: Add bar plot showing Male and Female user counts by country, include top 10:

TODO: Choropleth map showing selected country(s).

Here are a few screen shots:

Here is the code:

Here is a link to Plotly Studio dashboard:

.