EDA with Categorical Variables Lab

Resources: https://github.com/fenago/cts245X/tree/main/EDA

Part 1: Analyzing a categorical variable

Open the Power BI report named `3_1_single_categorical.pbix` and load the CSV file `glassdoor.csv` into the report.

Create a table showing the distinct count of reviews per `JobRole` .

Visualize the counts per `JobRole` with a new *Clustered bar chart.*

Update the table to add `MonthlyIncome` as an average. Sort so the highest average monthly income is in the first row.

Visualize the average monthly income by `JobRole` using a *Clustered bar chart.* The highest average monthly income should be on the top.

Add a constant line for the average monthly income. Make sure to format the label so it is readable.

How many job roles are above the overall average, represented by the constant line?

Part 2: Proportions with categorical variables

On a new page named "proportions", create a matrix of `JobRole` by `MaritalStatus` showing the distinct count of reviews.

Create a *100% Stacked column chart* of the percentage of reviews by `JobRole` and `MaritalStatus` .

Add data labels and round each value to the nearest whole number.

Which job role has the greatest proportion of single invididuals?

Part 3: Analyzing multiple categorical variables

On a new page named "multiple", create a *Matrix* showing the distinct counts of reviews by `Gender` and `Education`.

Duplicate the matrix and transform to a *100% Stacked column chart* to visualize the counts among these two categorical variables. Use `Gender` variable in the *Legend*.

Create a new matrix showing the average `HourlyRate` by `Gender` and `Education`.

Duplicate this matrix and transform to a *Clustered column chart*. Use `Gender` variable in the *Legend*.

Which educational level has the highest percentage of females and is the only educational level where females earn a higher average hourly rate than males?

Part 4: Creating a box plot

Add a new page and call it `box plot`.

Create a new column called `CareerStage` using the `SWITCH()` function. If `TotalWorkingYears` is less than or equal to 10, set the new variable to "Early"; if less than or equal to 25, set to "Mid", and if greater than 25, set to "Tenured". Otherwise, set the new variable to "NA".

https://learn.microsoft.com/en-us/dax/switch-function-dax

Create a new box plot showing the distribution of `MonthlyIncome` by `CareerStage`, using `ReviewId` as the aggregation level.

Duplicate the box plot and add `Gender` so that it compares males and females within each `CareerStage`.

Arrange all of the charts on the page to be readable. Feel free to close the *Visualizations* and *Fields* panes to help with sizing.

What is the average monthly income for "mid-career" professionals?

In which stage do female professionals tend to make more on average compared to male?