CSE 578 - Data Visualization
Project milestone 2
Course Project Final Report

Kedar Sai Nadh Reddy Kanchi
School of Computing and Augmented Intelligence
Arizona State University, Tempe
kkanchi@asu.edu

Problem Statement

UVW College, a local college is looking to bolster enrollment and for this it has chosen salary as a key demographic to determine criteria for marketing its degree programs. Therefore, for UVW College to get it's results, there are three things to be done:

- To develop marketing profiles of individuals using data supplied by the United States Census Bureau, with a focus on $50,000 as a key number for salary.
- To identify the factors that determine the individual's income.
  - There are many key variables that must be assessed for individuals making less than and more than $50,000, including age, gender, education status, marital status, occupation, etc.
- To develop an application to predict the income of an individual based on different values of the input parameters so that UVW College can tailor their marketing efforts when reaching out to the individuals.

Goals and a business objective

Goals

The primary goal of this project is to identify patterns in the data set by plotting visualizations to assist in determining the factors that account for determining an individual's income and present them to UVW executives. Following that, based on the preliminary analysis, the secondary goal of this project is to create machine learning models that accurately predict people's income.

Business Objective

The business objective of this project is for the UVW marketing team to use the machine learning model and analysis results to tailor their marketing efforts of reaching out to individuals.

Assumptions

- **Completeness and Comprehensiveness:** When developing a machine learning model, you must be cautious of missing data in the data-set. This is due to the fact that incomplete data is just as dangerous as inaccurate data. As a result, I assume that the data-set is incomplete and that there may be some gaps in the data collection, which may result in a distorted view of the overall picture. It is critical to understand the entire set of requirements that comprise a comprehensive set of data in order to determine whether or not the requirements are being met. As a result, I would be performing data pre-processing to address any gaps in data collection.
- **Data-set is accurate and precise:** I assume that the data-set provided to us is accurate and precise. It cannot have any erroneous elements and must convey the correct message without being misleading. Without understanding how the data will be consumed, ensuring accuracy and precision may be off-target or more expensive than necessary.
- **Feature Selection:** I believe that the features that have the greatest influence on class prediction will produce more interpretable patterns. As a result, a large number of visualizations and analyses are based on unbiased features.

User Stories

- **User Story #1:** As a data analyst at XYZ working for UVW, I'd like to know if an individual's age and capital-gain are relevant factors in determining their salary, so I can determine whether it should be included in my salary prediction model.
- **User Story #2:** As a data analyst at XYZ for UVW, I am asked to show hours-per-week and capital-gains for people earning more than $50,000 and less than $50,000 in order to better understand the demographics of target groups and create a social profile of target students.
- **User Story #3:** As a data analyst at XYZ for UVW, I am asked to show hours-per-week and age for people earning more than $50,000 and less than $50,000 in order to better understand the demographics of target groups and create a social profile of target students.
- **User Story #4:** As a data analyst at XYZ for UVW, I am asked to show age, education-num and capital-gain for people earning more than $50,000 and less than $50,000 in order to better understand the demographics of target groups and create a social profile of target students.
- **User Story #5:** As a member of the UVW marketing team, I'd like to know if an individual's occupation is a relevant factor in determining their salary, so I can determine whether it should be included in my salary prediction model.

- **User Story #6:** As a member of the UVW marketing team, I'd like to know if an individual's marital-status is a relevant factor in determining their salary, so I can determine whether it should be included in my salary prediction model.
- **User Story #7:** As a member of the UVW marketing team, I'd like to know if an individual's workclass is a relevant factor in determining their salary, so I can determine whether it should be included in my salary prediction model.
- **User Story #8:** As a data analyst at XYZ for UVW, I am asked to show occupation, education and marital-status for people earning more than $50,000 and less than $50,000 in order to better understand the demographics of target groups and create a social profile of target students.

Background Work Done

Loading Data

- The data-set comprising 32,561 data points were loaded into a pandas data-frame.
- A header row was added to the data-frame to contain feature names. These features are as follows: age, workclass, fnlwgt, education, education-num, marital-status, occupation, relationship, race, sex, capital-gain, capital-loss, hours-per-week, native-country, and class.

Understanding the Data Collection

- The last column is the result column, Class - which has two classes/categories - above 50k (">50k") and below 50k ("<=50k")
  - Out of these 32,561 individuals in the data-set, 24,720 earn a salary less than $50,000 while the remaining 7,841 individuals earn a salary greater than $50,000.
- Apart from the resultant column at the end, the data set has a total of 14 features. And out of these 14 features, 8 features are having categorical data.
- Now, coming to the data in the data-set, the data is in the form of a comma-separated values. Also, there are a total of 4,262 missing values which is represented by "?" in the data.

Data Pre-Processing/Cleaning

- I executed the command, "df.isna().sum().sum()", to check if there are other missing values other than the ones represented by "?" in the data.
  - But the command "df.isna().sum().sum()" resulted in a value of 0, which means that there are no other missing values in the whole data-set apart from the ones represented by "?".
- So, now the following step for me was to handle the missing values in the data-set represented by "?".
  - There were 4,262 instances of values represented by "?" out of the total 32,561 rows which constitutes to approximately 13% of the complete data.
  - Since, the rows containing the missing values, constitutes to a very less percentage, I have decided to remove these 4,262 rows from the data-set.
  - Therefore, for each column, I extracted all the rows that did not contain "?".
  - After removing the rows, that contained "?" values, the total number of rows in the cleaned data-set are 30,162. So, this means that effectively only 2,399 rows have been removed as opposed to 4,262 rows. So, the number of removed only constitute to 7% of the initial data-set.

Visualizations

Individual Feature Analysis

To carry out this uni-variate analysis, each of the continuous and categorical features were analyzed individually using data exploration techniques to identify patterns. Data visualization tools such as pie charts, bar charts, histograms, and box plots were used to accomplish this. To better comprehend the underlying distribution, other statistical techniques like mean, median, and standard deviation were applied when necessary.

This analysis helps to pinpoint any potential issues with the data, such as outliers. This process helps to guarantee the accuracy of the model that will be developed from the data.

As a result of the Individual Feature Analysis, I have reached the preliminary conclusion that the most important features are age, education-number, marital-status, occupation, and capital-gain.

In addition to the five features mentioned above, less significant or redundant features, I believe capital-loss, fnlwgt, and education are less significant or redundant features.

Therefore, such redundant features might not required for building the marketing profiles in the future alongside the five mentioned above.

Multivariate Feature Analysis

After completing the individual feature analysis, I strengthened my initial analysis with multivariate feature analysis tools such as the mosaic plot, parallel coordinate plot, and scatter plot. I was able to identify a number of key features as a result of this. This served as the foundation for the machine learning analysis, which included feature engineering of the key features.

Initial Analysis of Individual Features

Age

As the age of the person increases, the salary of the person decreases. People in the range of 30-55 have the best probability in getting salary above 50k.

fnlwgt

As illustrated in the figure-1, fnlwgt has statistical properties that are similar for both classes of data. Their distribution is also similar, as shown in the figure-1. As a result, this feature may be ineffective at distinguishing the two types of data.

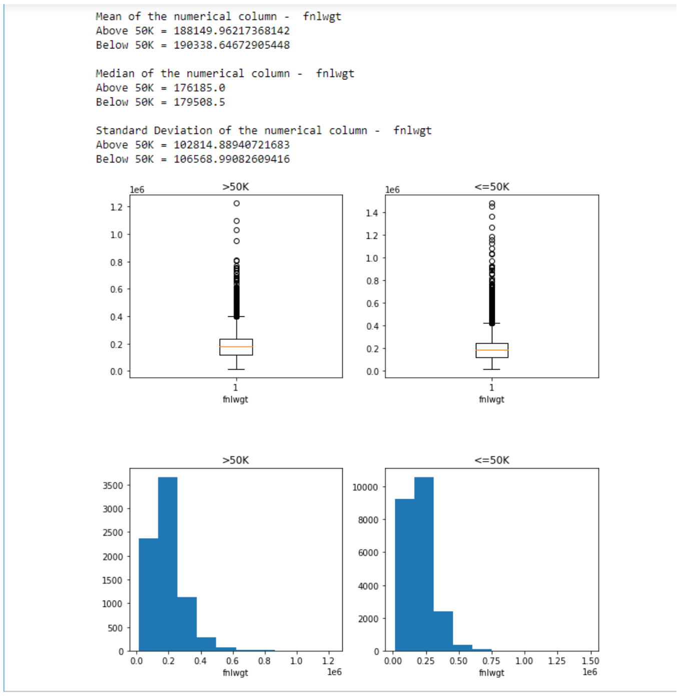

Figure 1: Numerical Analysis for the fnlwgt Feature

education-num

The graphs in Figure-2 clearly show that people with an education level of 10 or higher have a better chance of earning more than 50k. When viewed alongside with the education feature, this data shows that people with bachelor's degrees and higher have a better chance of earning more than 50k.

Capital-gain

The graphs clearly show that a person is more likely to earn more than 50k if the capital gain is high.

Capital-loss

As illustrated in the figure-3, capital-loss is more or less the same for both the classes of people. Their distribution is also similar, as shown in the figure-3. As a result, this feature may be ineffective at distinguishing the two types of data.

hours-per-week

The graphs clearly show that those who work more than or equal to 40 hours per week fall inside the ">50K" category.

Workclass

The data representation for the workclass feature is very stable across both classes.

Education

From the analysis made for the education-num feature, it is evident from the graphs as you can see in figures 4 and 5 that the people with education more than bachelors have a higher chance of getting salary more than 50k. Because this same information is encoded in the education-num feature, I have decided to exclude the education feature from my model.

Marital status

85 percent of people earning more than $50,000 have a marital status of Married-civ-spouse. With a 6 percentage each, the next highest percentage of people earning more than

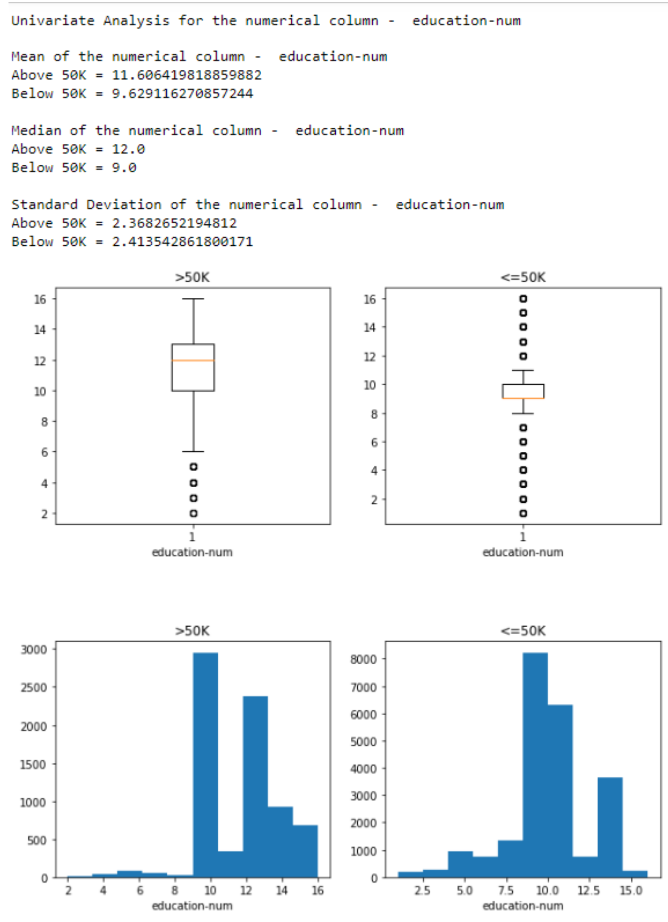

Figure 2: Numerical Analysis for the education-num Feature

$50,000 have a marital status of divorced or never-married. People with the marital status of never-married, on the other hand, make up the greatest proportion of those earning less than or equal to $50,000, accounting for 41 percent of the total.

Occupation

According to the graphs, people who work as Exec-managers or Prof-specialists account for a combined 50% of those earning more than $50,000. While the distribution of people earning less than $50,000 is close, occupations such as Craft-Repair, other-services, and Adm-Clerical take the top spot with 14% each. The graphs also show that people with occupations such as Craft-Repair, Sales, Tech-Support, and Protective-Services are evenly distributed between the two classes.

Relationship

This particular feature has a greater influence on salary, as people in the relationship "Husband" account for 75% of all people earning more than $50,000. Surprisingly, people in the relationship "Husband" have the highest distribution, accounting for 30% of all people earning less than or equal to $50,000. Similar to the previous inference based on the "Marital-status" feature, which showed that unmarried

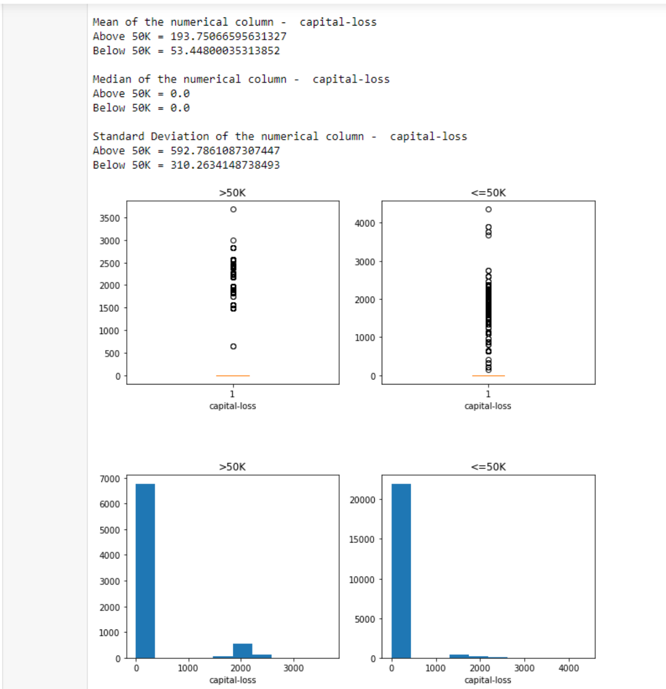

Figure 3: Numerical Analysis for the capital-loss Feature

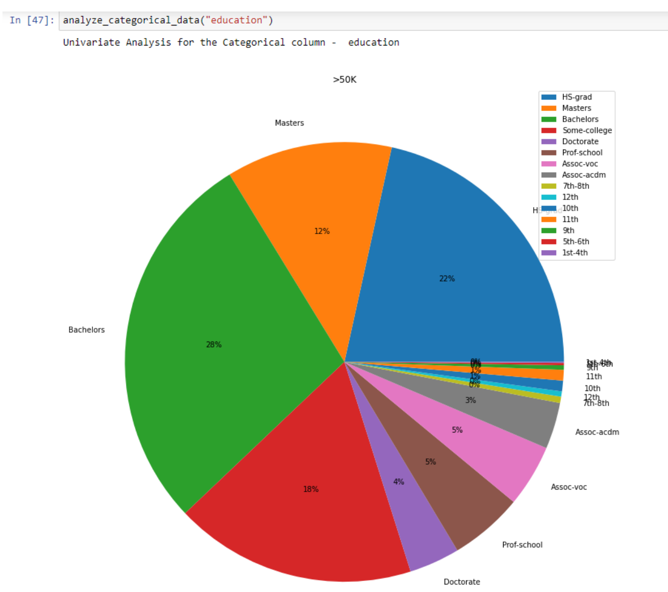

Figure 4: Categorical Analysis for the Education Feature for >50k class

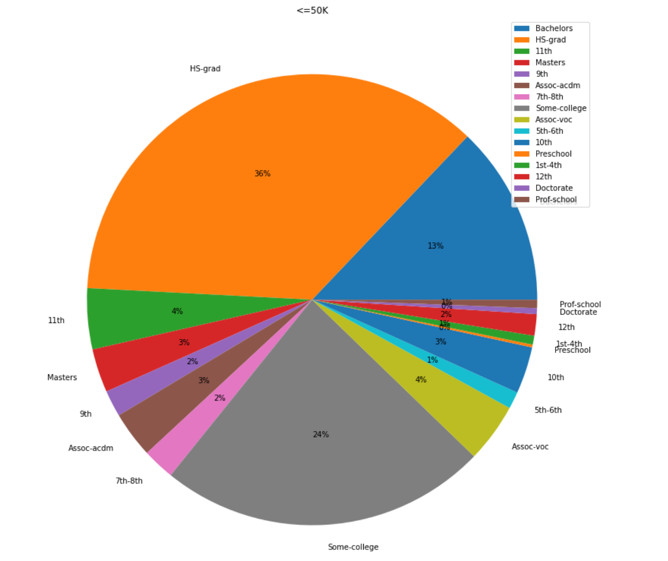

Figure 5: Categorical Analysis for the Education Feature for <=50k class

or never married people are more likely to earn less than \$50,000, people in the "Not-in-Family" category and the "Unmarried" category account for 43% of the total people earning less than or equal to \$50,000.

Race

This feature entirely benefits "White" people in both classes. The graphs displayed in figures 6 and 7, clearly show that 91% of those earning more than \$50,000 are of the "White" race, while 84% are of the "White" race. The "Black" race people are a miles away, accounting for 5% and 11% of the total. As a result, the "Race" attribute has little impact on people's salary ranges, so I've decided to exclude it from my model.

Sex

The graphs clearly show that the male sex dominates the class of people earning more than \$50,000, accounting for 85% of the total. Whereas the male sex has a slight advantage, accounting for only 62% of those earning less than or equal to \$50,000. Despite the fact that male sex dominates in both classes, this feature may play a role in classifying people as earning less than or equal to \$50,000.

Native-Country

The distribution of this feature is very similar to the distribution of the Race feature. This is due to the fact that the distribution for the native-country feature clearly shows that people from the United States clearly dominate in both classes, accounting for 93% in the >50k class and 91% in the <=50k class. Nationalities such as Jamaica, Mexico, Puerto Rico, Honduras, Colombia, Ecuador, Laos, Haiti, Portugal, Dominican Republic, El-Salvador, Guatemala, Peru, Outlying-US, Trinidad & Tobago, Nicaragua, and Vietnam are more likely to earn less than \$50,000.

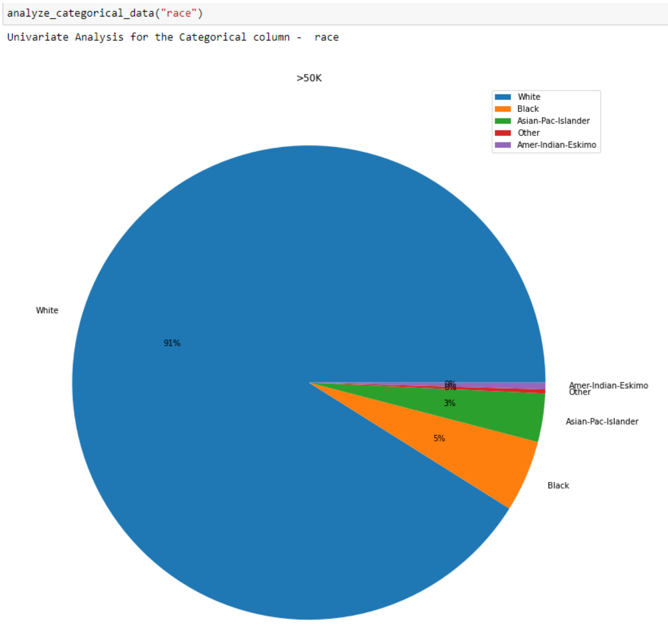

Figure 6: Categorical Analysis for the Race Feature for >50k class

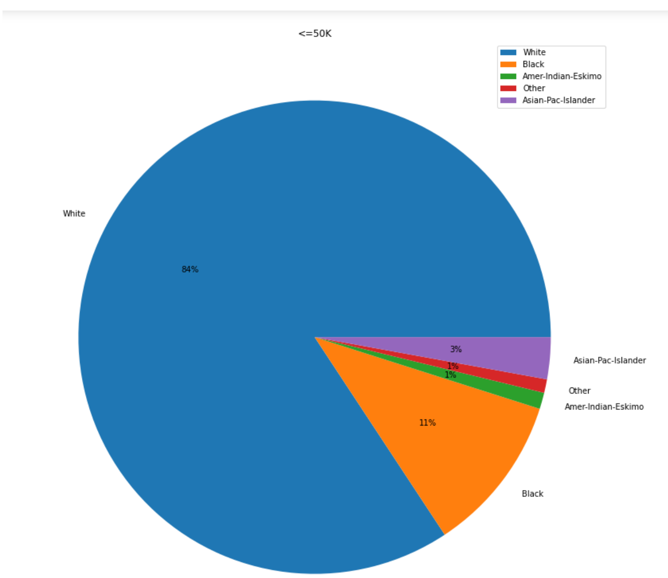

Figure 7: Categorical Analysis for the Race Feature for <=50k class

Important Features

As previously stated, based on my preliminary findings, I identified age, capital gain, occupation, marital status, and education number as the most important factors influencing the class variable, i.e. an individual's salary range. As a result, I used multivariate analysis to plot some graphs between the aforementioned features or against the class feature to gain more insight.

The following are the key graphs that helped me identify the key features in the dataset:

User Story #1 - Features Covered : Capital-Gain and Age

I used the following code to get the visualization below.

```
def plot_scatter_plot_diff(column1, column2):
    plt.close()
    fig, axes = plt.subplots(ncols = 1, nrows = 2, figsize = (30 , 30))
    fig.subplots_adjust(hspace = .5)

    colors = df['class']
    x = df[column1]
    y = df[column2]
    axes[0].scatter(x , y , c = colors)
    axes[0].set_title("purple <=50K, yellow >50K")
    axes[0].set_xlabel(column1)
    axes[0].set_ylabel(column2)
```

Figure 8: Scatter plot Code: Capital-gain vs Age

The above code gives me the following scatter plot:

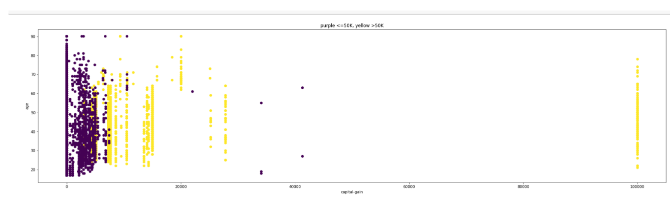

Figure 9: Scatter plot: Capital-gain vs Age: For both the classes

Inferences:

- With the exception of a few outliers, it is clearly evident that there is a clear distinction between the two types of data.
- Furthermore, as previously stated, the graphs clearly show that individuals with high capital gains are more likely to earn more than $50,000 per year.

User Story #2 - Features Covered : Capital-Gain and hours-per-week

The code shown in Figure-8 gives me the following scatter plot as shown in Figure-10 between the features Capital-Gain and hours-per-week:

Inferences:

- If a person's capital gain is low, he or she is more likely to fall into the category of people earning less than or equal to $50,000, regardless of the number of hours worked per week.

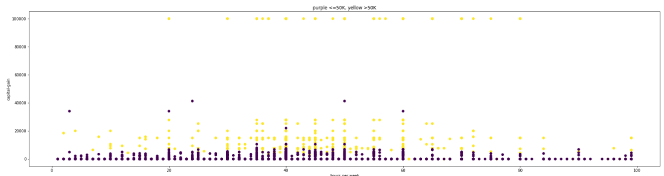

Figure 10: Scatter plot: Capital-gain vs hours-per-week: For both the classes

- Similarly, If a person's capital gain is very high, he or she is more likely to fall into the category of people earning more than $50,000, regardless of the number of hours worked per week.
- However, if a person's capital gains are moderate, he or she is more likely to fall into the category of people earning more than $50,000 if they work 40 to 60 hours per week.

User Story #3 - Features Covered : Age and hours-per-week

The code shown in Figure-8 gives me the following scatter plot as shown in Figure-11 between the features age and hours-per-week:

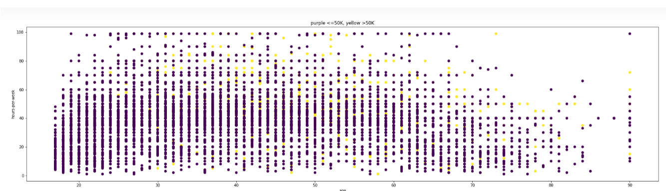

Figure 11: Scatter plot: age vs hours-per-week: For both the classes

Inferences:

- If a person is under the age of 35, he or she is more likely to earn less than or equal to $50,000, regardless of the number of hours worked per week.
- If a person is between the ages of 35 and 60, he or she is more likely to earn more than $50,000 if they work a minimum of 40 hours per week.
- As a person approaches 60, he or she is unlikely to earn more than $50,000. According to the graph, there are a few instances where people over the age of 60 fall into the category of people earning more than $50,000 despite working short hours, but as age increases, those people only fall into this category when they work more than 45 hours per week. When people reach the age of 80, the proportion of people earning more than $50,000 begins to fall dramatically.
  - In conclusion, the scatter plot indicates that your salary increases as you work longer, so eventually older people may no longer fall into the category of those making less than $50,000. This demonstrates that while an individual's starting salary may be less than $50,000 on average, as they age and gain experience, there is a nearly certain likelihood that they will earn more than $50,000.

User Story #4 - Features Covered : Capital-Gain, Age, and education-num

To analyze multivariate numerical data, a parallel coordinates plot is used. Each row in the data table is represented as a line or profile in a parallel coordinate plot. Each row attribute is represented by a point on the line. It enables you to compare samples or observations across multiple numerical variables. Each attribute or variable has its own axis. All of the axes are evenly spaced and parallel to one another. Parallel coordinate plots resemble line charts in appearance, but the way data is translated into a plot is vastly different.

I used the following code as you can see in Figure-10 to get the parallel coordinates plot visualization shown in Figure-11.

- In the code, I had to sample 30 records at random.
- This is due to the fact that the number of records that fall into the <=50k category is significantly greater than the number of records that fall into the >50k category.
- As a result, the lines representing the <=50k class were very dominant in the plot, and I had to reduce the number of records representing the <=50k class. That is why I chose a sample of 30 records at random.
- Finally, each of the features are scaled down to value between 0 and 1 to get a better plot.

```
frame_pc = df[['education-num', 'age', 'capital-gain', 'class']].copy()
frame_np_array = MinMaxScaler().fit_transform(frame_pc.values)
frame_pc = pd.DataFrame(frame_np_array)
df.index = frame_pc.index
frame_pc['salary-range'] = df['salary-range']
frame_pc.columns = ['capital-gain', 'education-num', 'age', 'class', 'salary-range']
frame_pc_below_50K = frame_pc[frame_pc["class"] == 0.0].sample(n=30)
frame_pc_above_50K = frame_pc[frame_pc["class"] == 1.0].sample(n=30)
frame_pc = pd.concat([frame_pc_below_50K, frame_pc_above_50K])
pd.plotting.parallel_coordinates(frame_pc, 'salary-range', cols=['education-num', 'age', 'capital-gain'], color=('#FF0000',
                    '#FFD700'))
plt.show()
```

Figure 12: Parallel Coordinate Plot Code: Capital-gain vs Age vs Education-num

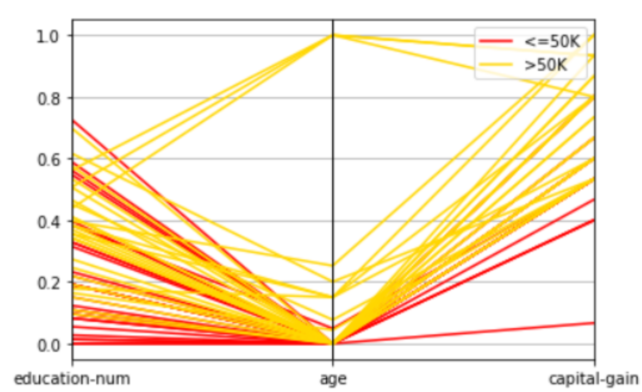

Figure 13: Parallel Coordinate Plot: Capital-gain vs Age vs Education-num: against both the classes

Inferences:

- The parallel coordinate plot demonstrates that combining these three features allows us to distinguish between the yellow and red lines.

- The box plot in Figure-2 for the education-num feature demonstrates that the distribution of education between the two classes of data varies greatly.
- The same box plot in Figure-2 for the education-num feature clearly shows that people with a high education number are more likely to earn more than $50,000.
- When I combine the readings from the parallel coordinate plot with the distributions for the age feature versus the class feature, I can confidently state that people between the ages of 30 and 60 earn more than younger people.
- Since experience pays more money and age directly correlates to experience, this makes intuitive sense to the majority of people. UVW College can therefore easily devise a sales strategy that persuades people that if they have a higher degree than their current one, they have a higher chance of earning more than $50,000.

User Story #5 - Features Covered : Occupation

Analysis between Occupation and Salary-Range

```
plt.close()
fig, axes = plt.subplots(ncols = 1, nrows = 1, figsize=(20 , 8))
fig.subplots_adjust(hspace = .5)
mosaic(df, ['occupation', 'salary-range'], ax = axes, axes_label = True)
plt.show()
```

Figure 14: Mosaic Code: Occupation vs Salary

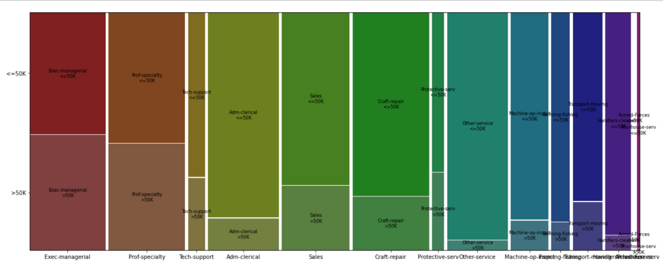

Figure 15: Mosaic Plot: Occupation vs Salary

Categories in the order as they appear: Exec-managerial, Prof-specialty, Tech-support, Adm-clerical, Sales, Craft-repair, Protective-serv, Other-service, Machine-op-inspct, Farming-fishing, Transport-moving, Handlers-cleaners, Armed-Forces, and Priv-house-serv.

Inferences:

- Similar to the occupation feature, for the marital-status feature in most categorical data, the distribution of the two classes is highly skewed, implying that this feature can be used to distinguish between the two classes.
- The graphs show that most people in the high specialization job categories on the left of the spectrum, such as Exec-managerial, Prof-specialty, Tech-support, and Sales, elicit a salary of more than $50,000. On the other hand, the majority of people employed in handmanship and craftsmanship jobs like farming, fishing, the armed forces, and craft repair generally earn less than $50,000 annually.
  - As a result, UVW College can take advantage of this to market degrees like business studies and computer science since the average salary in these fields is higher than $50,000.

User Story #6 - Features Covered : marital-status

```
plt.close()
fig, axes = plt.subplots(ncols = 1, nrows = 1, figsize=(20 , 8))
fig.subplots_adjust(hspace = .5)
mosaic(df, ['marital-status', 'salary-range'], ax = axes, axes_label = True)
plt.show()
```

Figure 16: Mosaic Code: marital status vs Salary

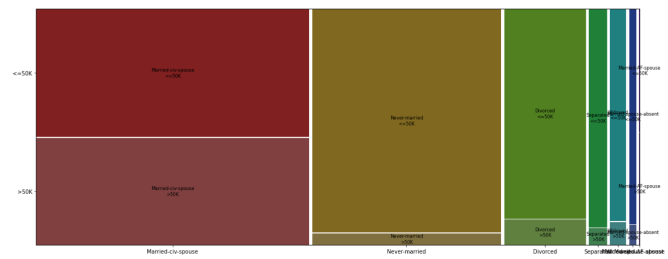

Figure 17: Mosaic Plot: marital status vs Salary

Categories in the order as they appear: Married-civ-spouse, Never-married, Divorced, Separated, Widowed, Married-spouse-absent, and Married-AF-spouse.

Inferences:

- In most categorical data, the distribution of the two classes is highly skewed, implying that this feature can be used to distinguish between the two classes.
- According to the graphs, people with a marital-status of “married-civ-spouse” are more likely to earn more than $50,000.
- According to the graphs, people with a marital-status of “never-married” are more likely to earn less than or equal to $50,000.

User Story #7 - Features Covered : workclass

Analysis between Workclass and Salary-Range

```
plt.close()
fig, axes = plt.subplots(ncols = 1, nrows = 1, figsize = (20 , 8))
fig.subplots_adjust(hspace = .5)
mosaic(df, ['workclass', 'salary-range'], ax = axes, axes_label = True)
plt.show()
```

Figure 18: Mosaic Code: Workclass vs Salary

Categories in the order as they appear: Self-emp-not-inc, Private, State-gov, Local-gov, Federal-gov, Self-emp-inc, and Without-pay.

Inferences:

- In most categorical data, the distribution of the two classes is stable with very less skewness, implying that this feature cannot be used to distinguish between the two classes.

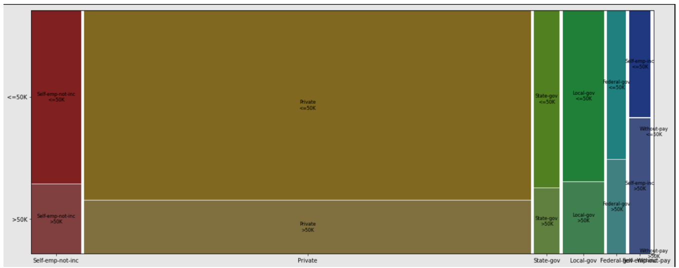

Figure 19: Mosaic Plot: Workclass vs Salary

- According to the graphs, people with a workcalss of “Private” are more likely to earn more than $50,000. -

User Story #8 - Features Covered : occupation, education and marital-status

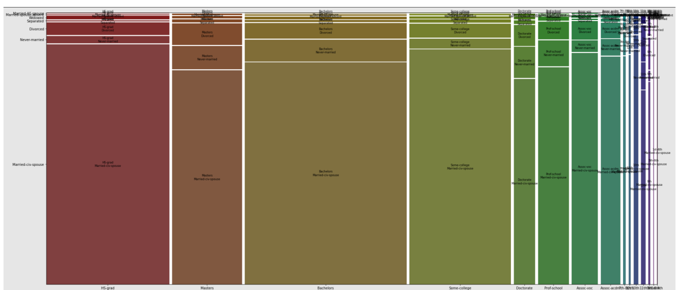

Figure 20: Mosaic Plot: education vs marital status vs Salary(for people earning more than $50,000)

Inferences:

- According to the graph in Figure-20, if a person is married, he or she is more likely to fall into the category of people earning more than $50,000, regardless of education.
- Again, from the same graph in Figure-20 it is clear that people with a marital-status of “married-civ-spouse” and an education higher than a bachelor’s degree have a very good chance of earning more than $50,000.
- According to the graph in Figure-21, people having a education of bachelors or masters and higher working as a “Exec-mangerial” or “Prof-speciality” have the highest chance of earning more than %50,000.

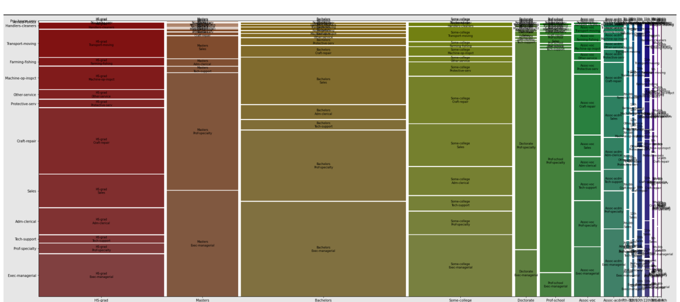

Figure 21: Mosaic Plot: education vs occupation vs Salary(for people earning more than $50,000)

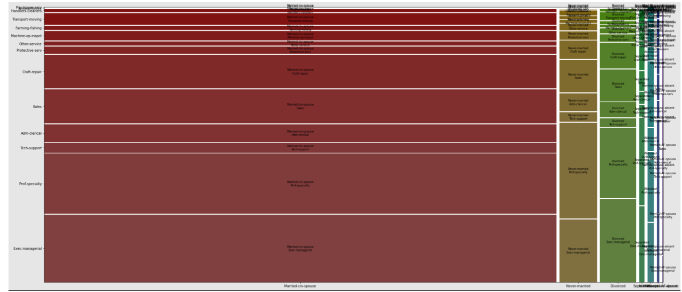

Figure 22: Mosaic Plot: marital status vs occupation vs Salary(for people earning more than $50,000)

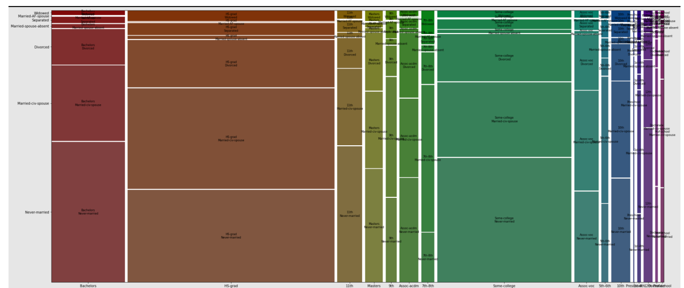

Figure 23: Mosaic Plot: education vs marital status vs Salary(for people earning less than or equal to $50,000)

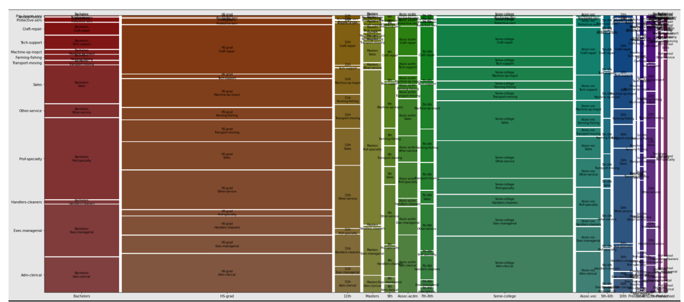

Figure 24: Mosaic Plot: education vs occupation vs Salary(for people earning less than or equal to $50,000)

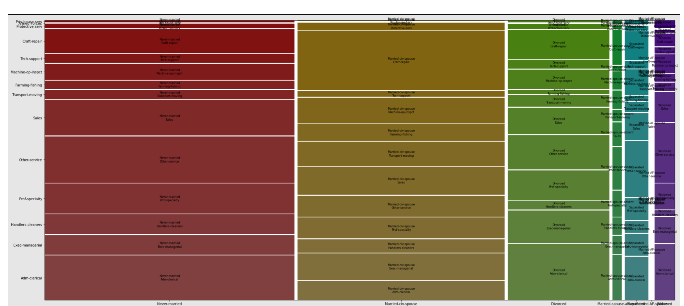

Figure 25: Mosaic Plot: occupation vs marital status vs Salary(for people earning less than or equal to $50,000)

- According to the graph in Figure-21, you can see that most categorical data, the distribution of the two features

```python
def plot_mosaic_class(column1, column2, column3):
    plt.close()
    cols = [column1, column2, column3]
    for i in range(3):
        for j in range(i+1, 3):
            print("Salary <=50K")
            fig, axes = plt.subplots(ncols = 1, nrows = 1, figsize = (33 , 15))
            fig.subplots_adjust(hspace = .5)
            mosaic(df_records_below_50K, [cols[i], cols[j]], ax = axes, axes_label = True)
            plt.show()
            print("Salary >50K")
            fig, axes = plt.subplots(ncols = 1, nrows = 1, figsize = (33 , 15))
            fig.subplots_adjust(hspace = .5)
            mosaic(df_records_above_50K, [cols[i], cols[j]], ax = axes, axes_label = True)
            plt.show()
```

Figure 26: education vs occupation vs marital status vs Salary

across the salary is highly skewed, implying that both these features are important and can be used to distinguish between the two income classes.

- According to the graph in Figure-22, people with a marital-status of "married-civ-spouse" dominate across all occupations, with the highest chance of earning more than $50,000, particularly those with an occupation of "Exec-mangerial" or "Prof-speciality". People with the occupations "Exec-mangerial" or "Prof-speciality" account for a higher percentage of all marital statuses.
- Looking at the graphs as show in Figures 20, 21, 22, 23, 24 and 25, I can say that an individual's salary will increase in direct proportion to the degree they have earned. UVW College's degree programs can draw students because a salary of more than $50,000 is essentially guaranteed. The dataset also includes a feature called education-num that tracks how long a person attended school for. The highest level of education can be used to estimate the approximate number of years of education completed; however, this feature is redundant and linearly dependent on education. The education-num feature can also be applied in this case to get the same result.

Questions Arised

- **Question-1:** The skewness of the data may make generating precise marketing profiles difficult. This is because 76% of the population earns less than $50000 per year. As a result, the question of whether the entire data-set should be included when creating marketing profiles arises.
  - **Solution:** The most effective way to address this issue is to sample the data so that there is an equal number of people earning less than and more than $50,000.
- **Question-2:** The data-set also revealed gender bias. This is because 67% of the people in the data-set are men and 33% are women. This bias may make marketing UVW College to female audiences difficult.
  - **Solution:** Again, this can be solved by selecting an equal number of men and women who fall into the less than and greater than $50,000 categories, allowing both genders to be effectively compared in terms of salary while also taking other important factors such as education-number and marital-status into account.
- **Question-3:** What types of data visualizations are suitable for conducting analysis of information?
  - **Solution for categorical data:** The pie chart and mosaic plots (multivariate) were employed to depict categorical data because they provide an accurate representation of this type of information.
  - **Solution for continuous data:** When it comes to continuous data, histograms, box plots, scatter plots, and parallel coordinate plots are frequently utilized for visualization purposes.

Not Doing

- **Machine Learning Model:** I am not currently creating a machine learning model for this task. However, I plan to explore the possibility of developing a machine learning analysis in the future.
- In the future, I plan to implement visual recommenders that will use a query builder to obtain the features of a new sample/user-data and then recommend the class label based on the features.
- Making class label predictions using range set data so that class labels can be predicted based on a range query. As an example, consider salary forecasting for people aged 35-40 who work 20-30 hours per week.
- Despite the fact that plots have been developed and UVW College has received advice on which demographic to target with each plot, a more thorough explanation of how to market this cannot be given at this time due to a lack of understanding of sales and marketing strategies. If you have more experience with sales and marketing strategies, you might also use this in the future.