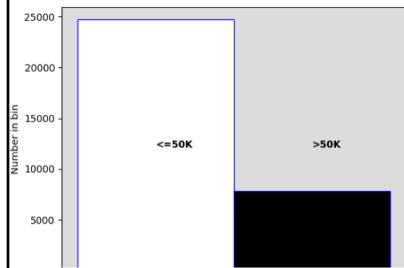


# The Top Six User Stories (Variables) Selected for Further Analysis

## An Executive Report by Team 4 (The Falcons)

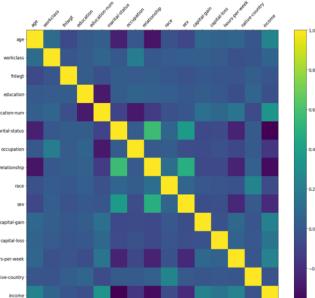


Presented before

The UVW Executives

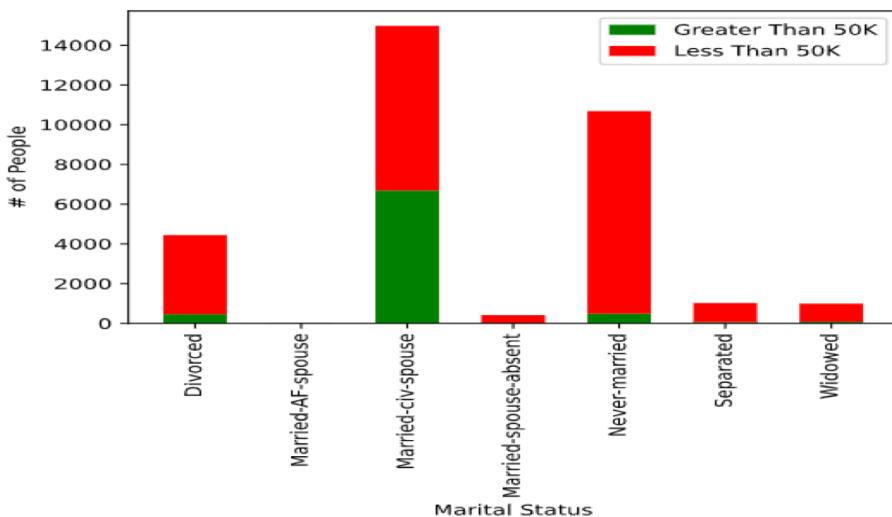
19<sup>th</sup> April, 2021

Correlation Matrix



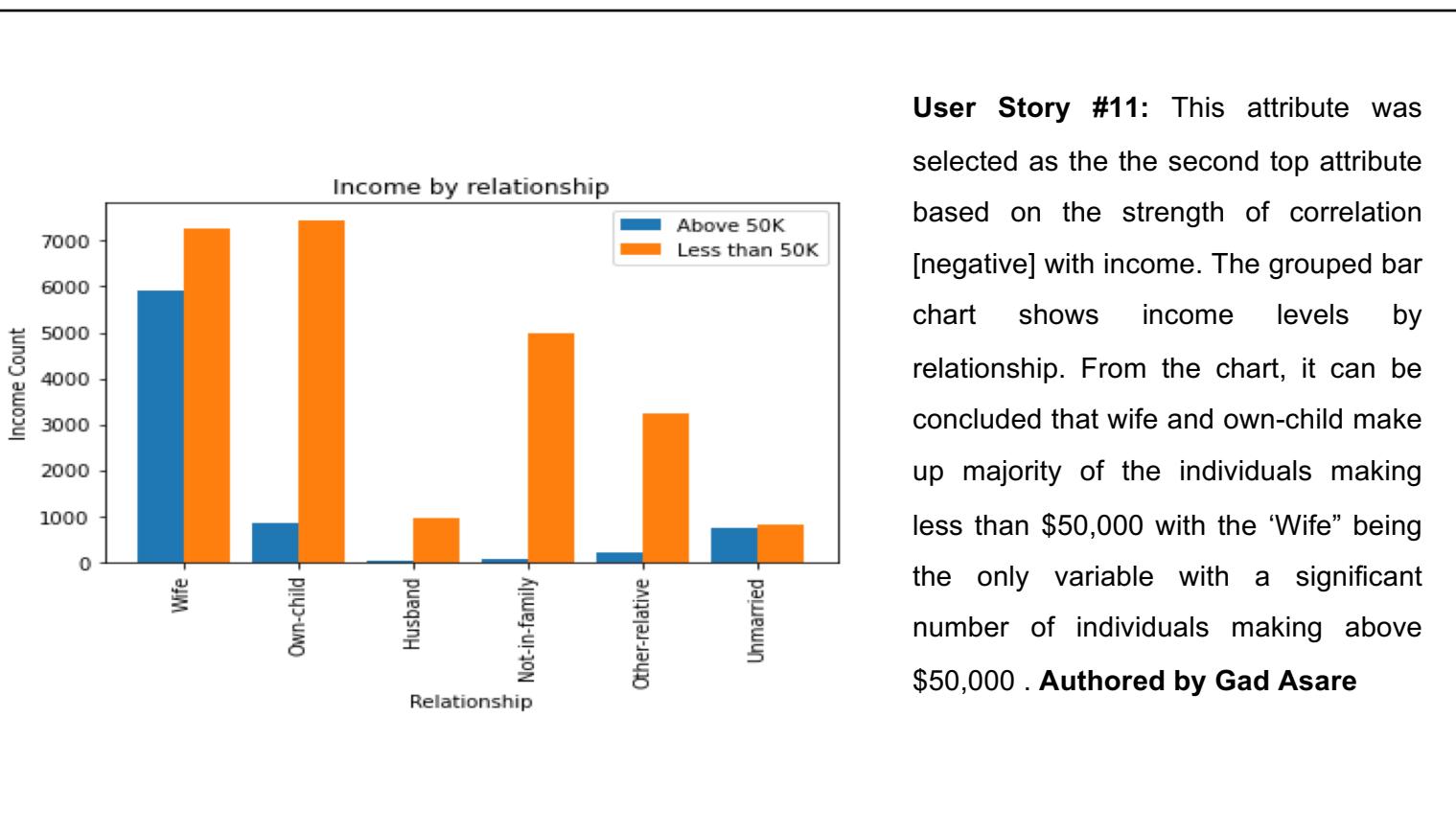
The aim of this project is to identify attributes that will increase enrollment in UVW College (measured by change in income status). The histogram shows the distribution of Income for all samples in the data set. It's a clear indication that the majority of people make less than or equal to 50k. The correlation matrix shows the strength of the relationship between the variables. We utilized below, the correlation coefficients (and fair representation of efforts) to identify our six top variables that will predict income level, in decreasing order of importance.

1

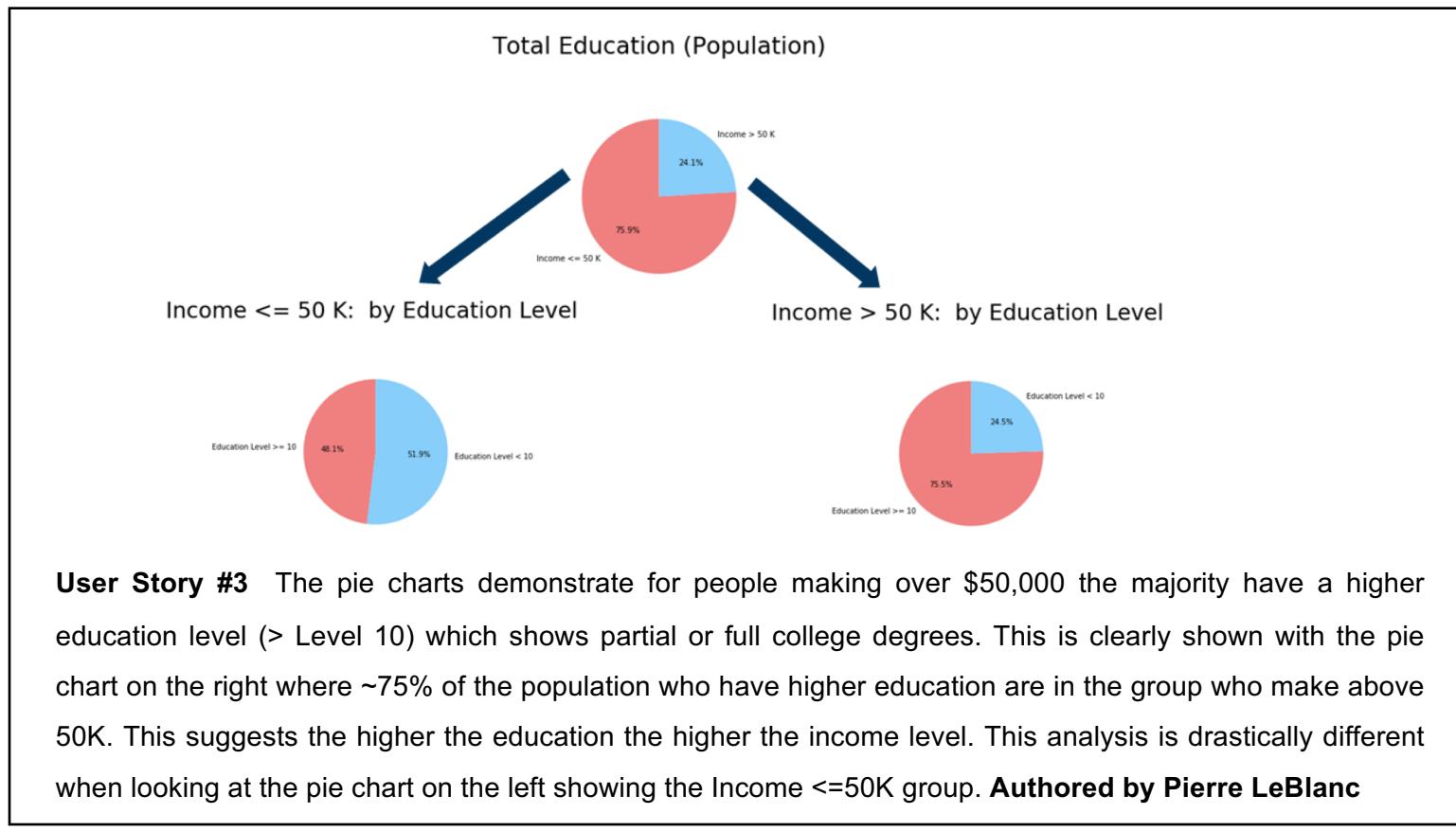


**User Story #7:** This attribute was selected as the top attribute based on the strongest correlation [negative] with income. The stacked bar chart shows what proportion of the people in each marital status category make more or less than the \$50,000 threshold. The only group that shows a significant portion making more than \$50,000 is the group of people who are Married. This would make a compelling choice for further evaluation as to how significant the difference of someone's marital status could be on their income level. **Authored by Intzar Singh**

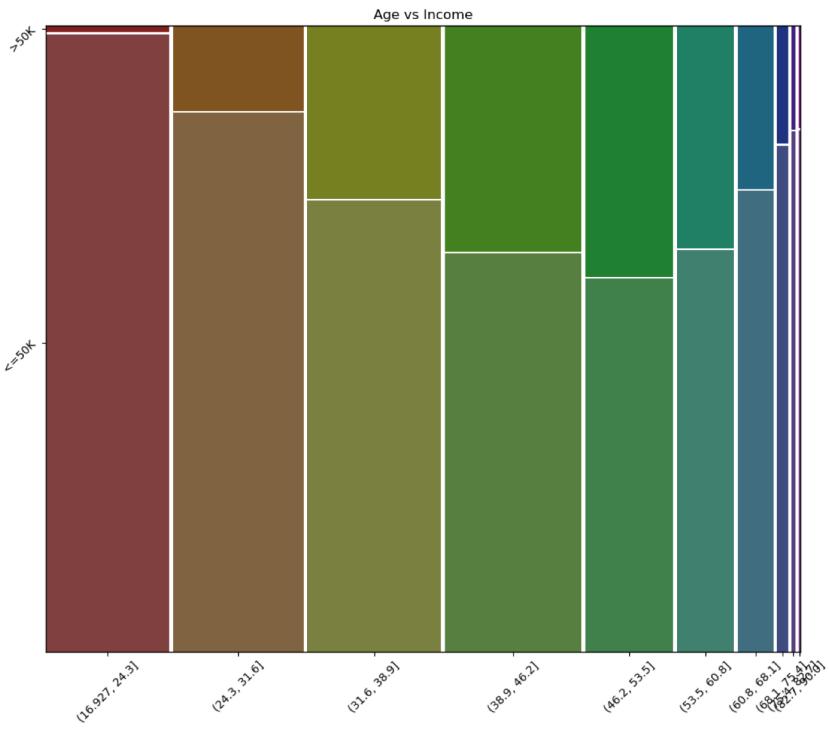
2



3



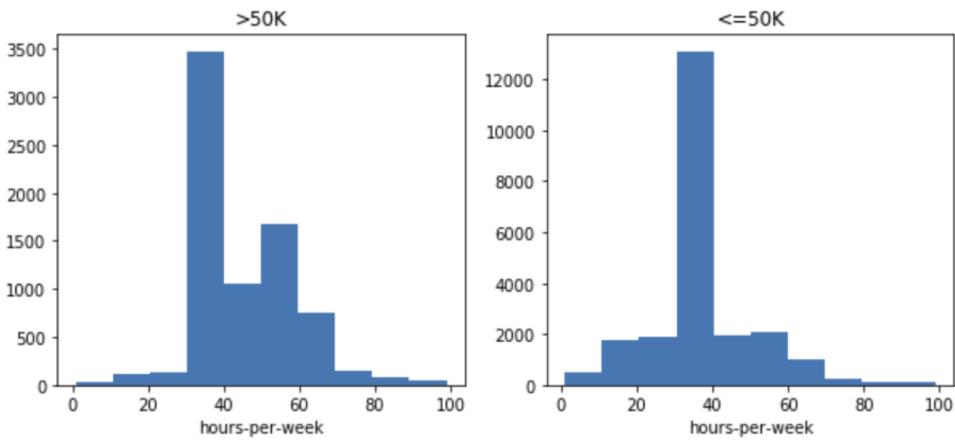
4



**User Story #1:** This mosaic plot shows Age to Income relationship. The width of each bar represents the distribution of data samples in each bin with respect to the total data pool. After age 46 the width of the bars rapidly deteriorate. With respect to age the distribution who make over \$50,000 is skewed to the right and this matches the observation from correlation matrix. This visualization shows that there will be merit in further analyzing the age of an individual and perhaps clustering age, income, and a third parameter to identify a deeper relationship to an individual's income. **Authored by Michael Salzarulo**

5

Distribution of Hours-Per-Week by Income



**User Story #6:** The histogram shows the most people working more than 40 hours a week belong to the ">50K" category. Also, in the >50k group it has less people who works less than 30 hours a week. This shows that people who spend more time working are likely to earn over \$50,000. This could be good variable for future use in the analysis.

**Authored by Jiteng Xu**

6

Distribution of Income by Sex



**User Story #12:** This sunburst plot shows that males are significantly more likely to earn more than 50K ( $p=0.0$  by chi square). Males dominate both income brackets. While males significantly makes up 85% (6662/7841) of people that earn greater than 50K, males only make up 61.2% (15128/24720) of people who make less than or equal to 50K. This analysis is the least correlated among the top six attributes. However it will make a good marketing strategy to both sex categories. To females, a motivation for crossing to a high higher earning. To males, ambition to maintain the higher earning bracket. **Authored by Hannah Ajoge**

7

## Conclusion:

- Our visualizations have identified six top variables that strongly correlate with income.
- We show that being married, a wife, possessing a college degree, younger than 46, working more hours and being a male is associated with earning more than \$50,000.
- Further analyzes of these top attributes will be necessary before implementation as part of our next marketing strategy.

THANK YOU FOR  
YOUR ATTENTION