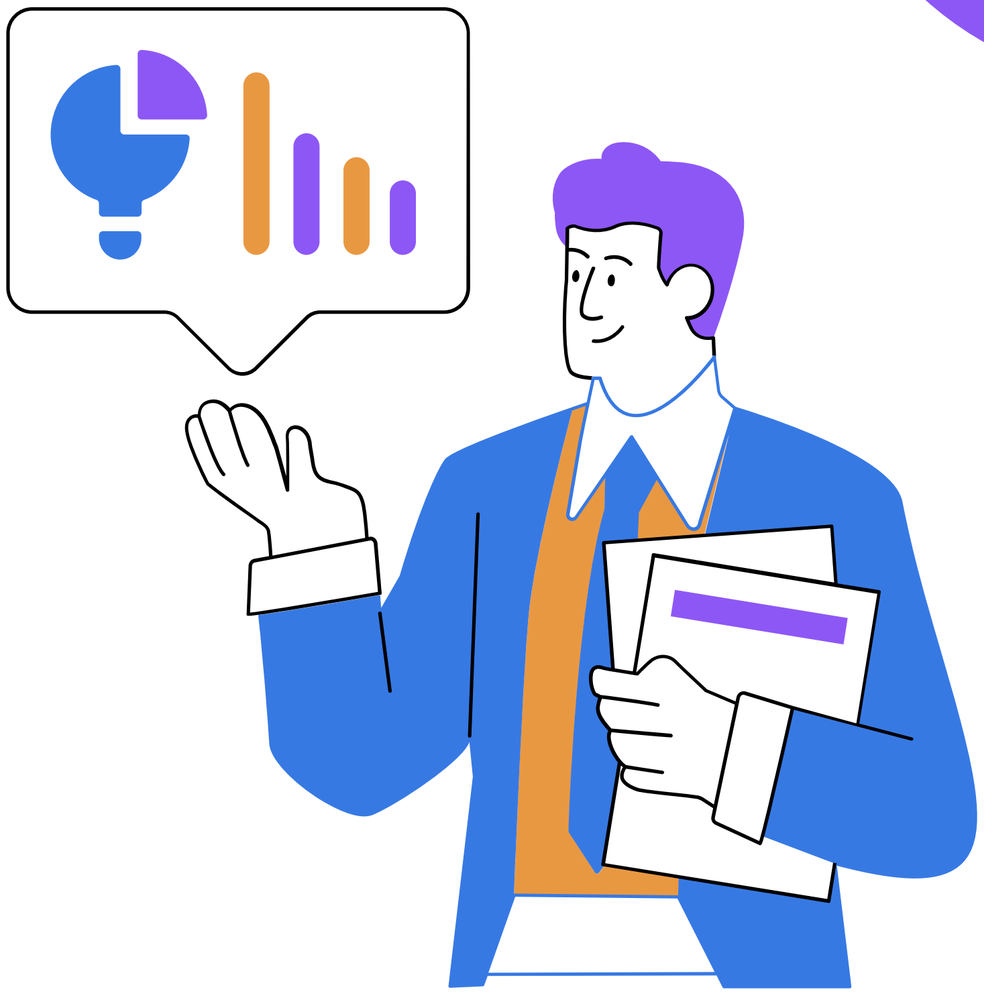
ANALYSIS OF CREDIT CARD DEFAULTERS

Marwan Ahmed - 202101214



Agenda

- Data Analysis
- Estimation Theory
- Confidence Intervals
- Hypothesis Testing/A-B Testing
- Regression Modeling

WHICH CUSTOMERS ARE AT HIGHEST RISK OF DEFAULT?

What is the overall distribution of key features?

Executives

Focus on trends and ROI from customer feedback.

Product managers

Use reviews to refine product features.

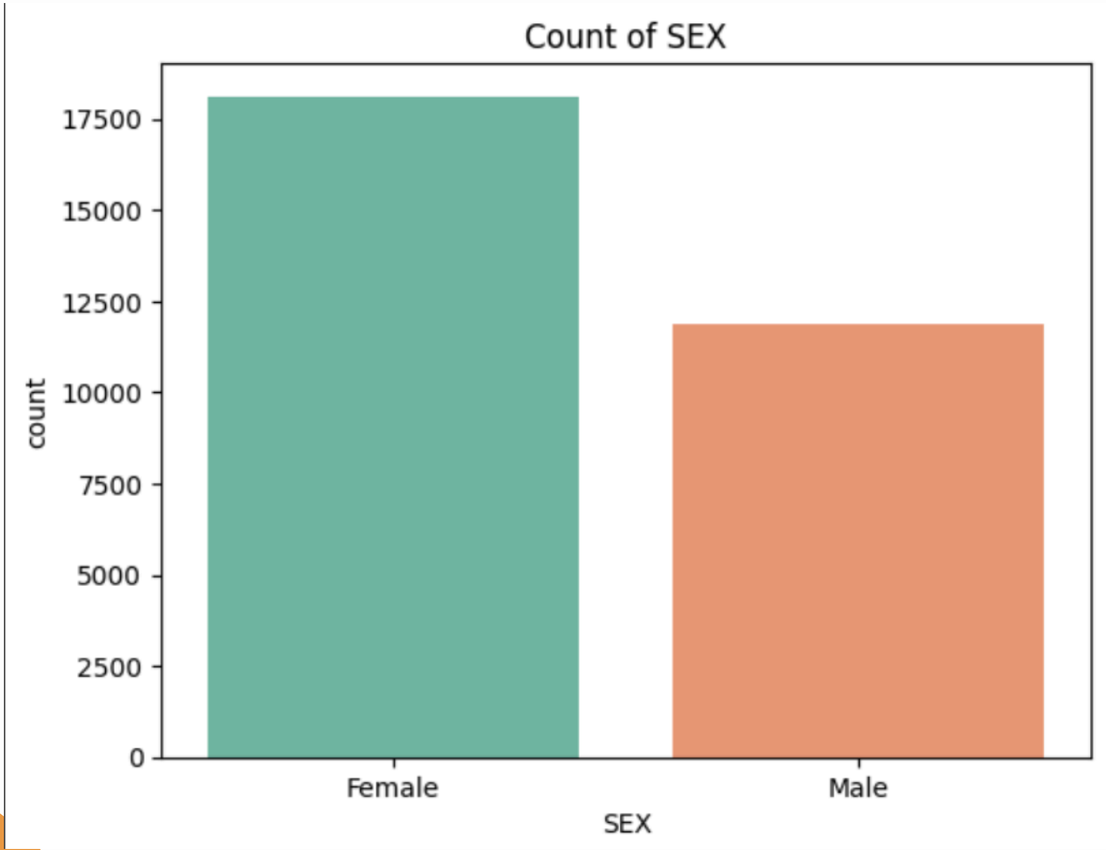
Marketing teams

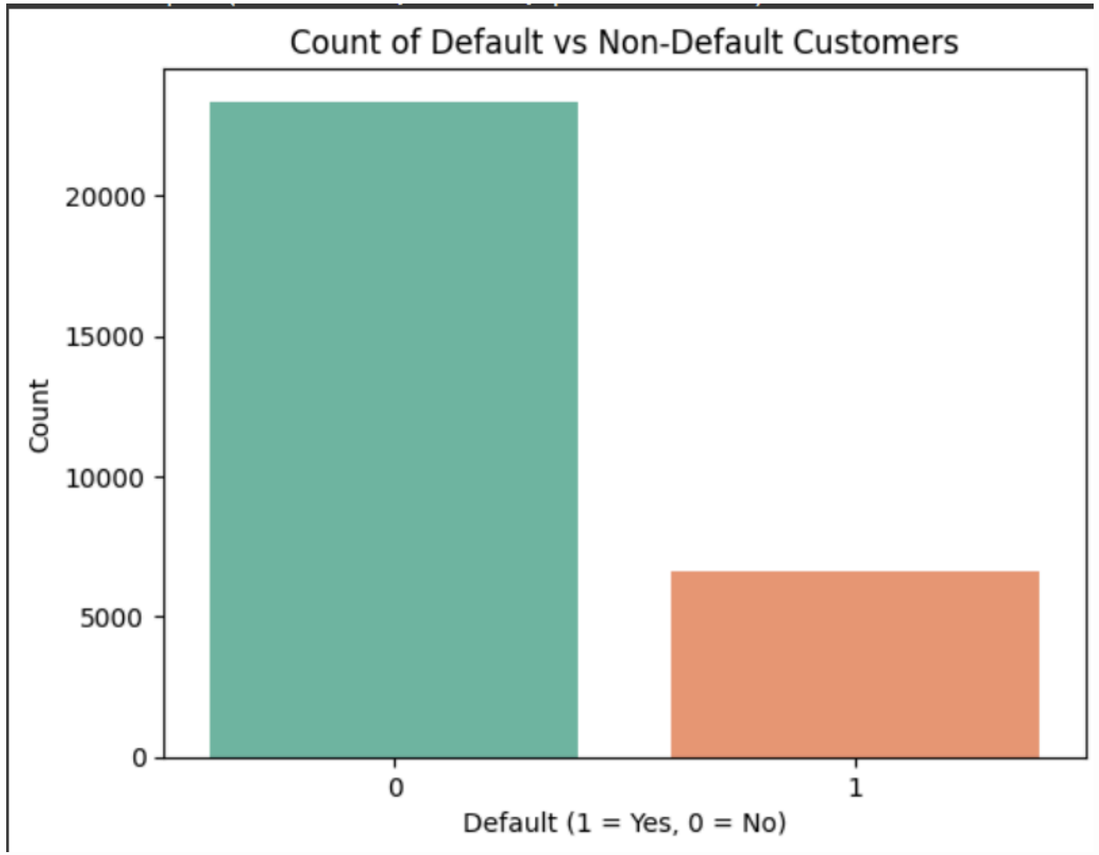
Monitor reputation and promote positive sentiment.

New Buyer

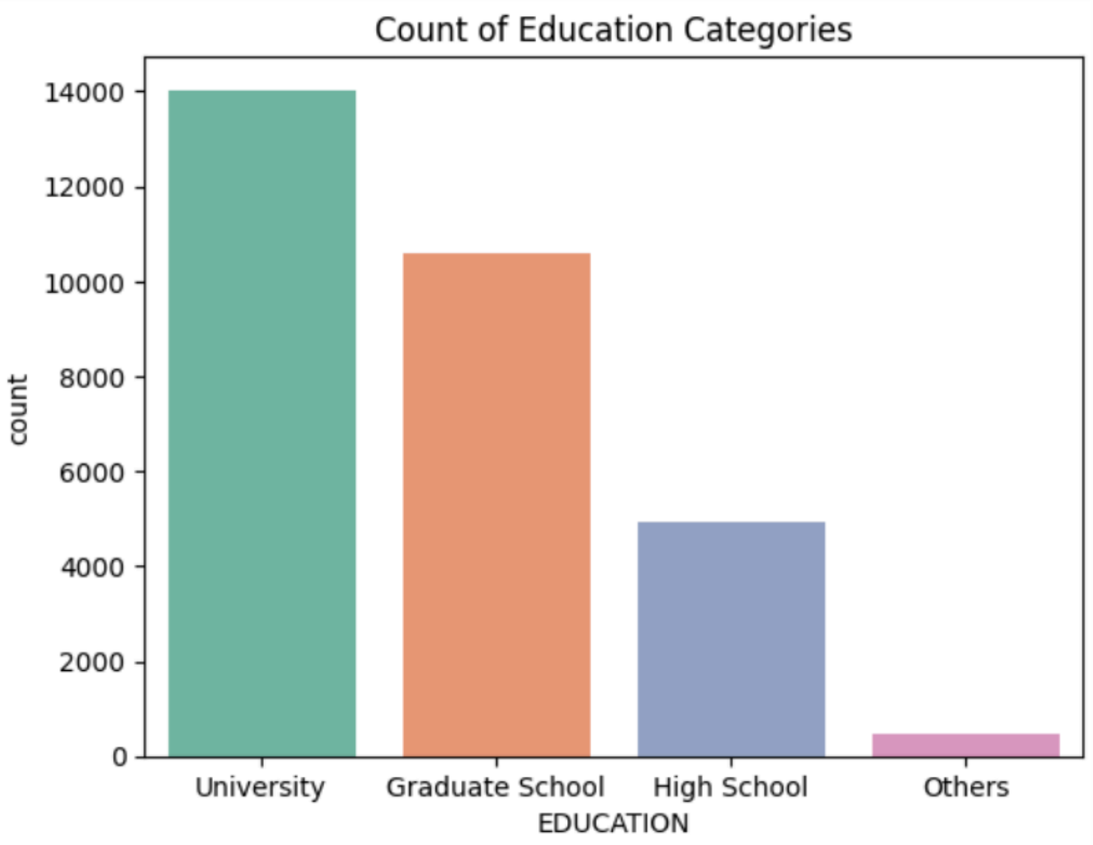
Relies on reviews before buying a product online

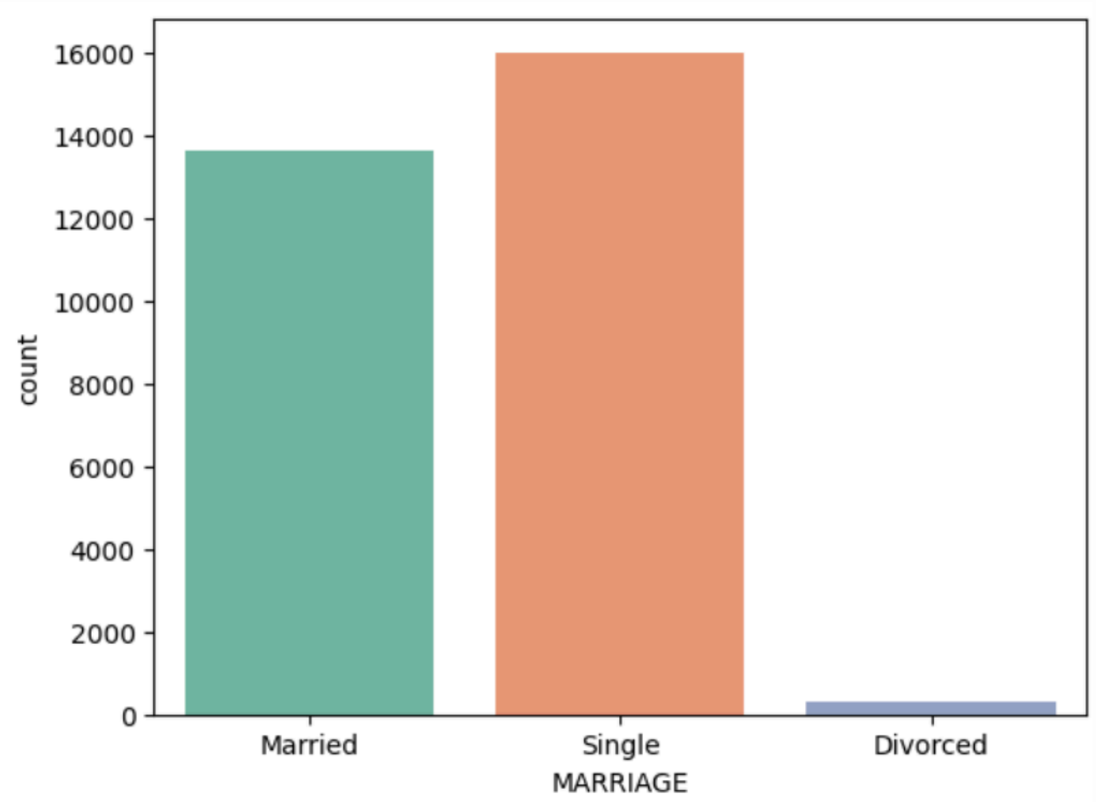
What is the overall distribution of key features?

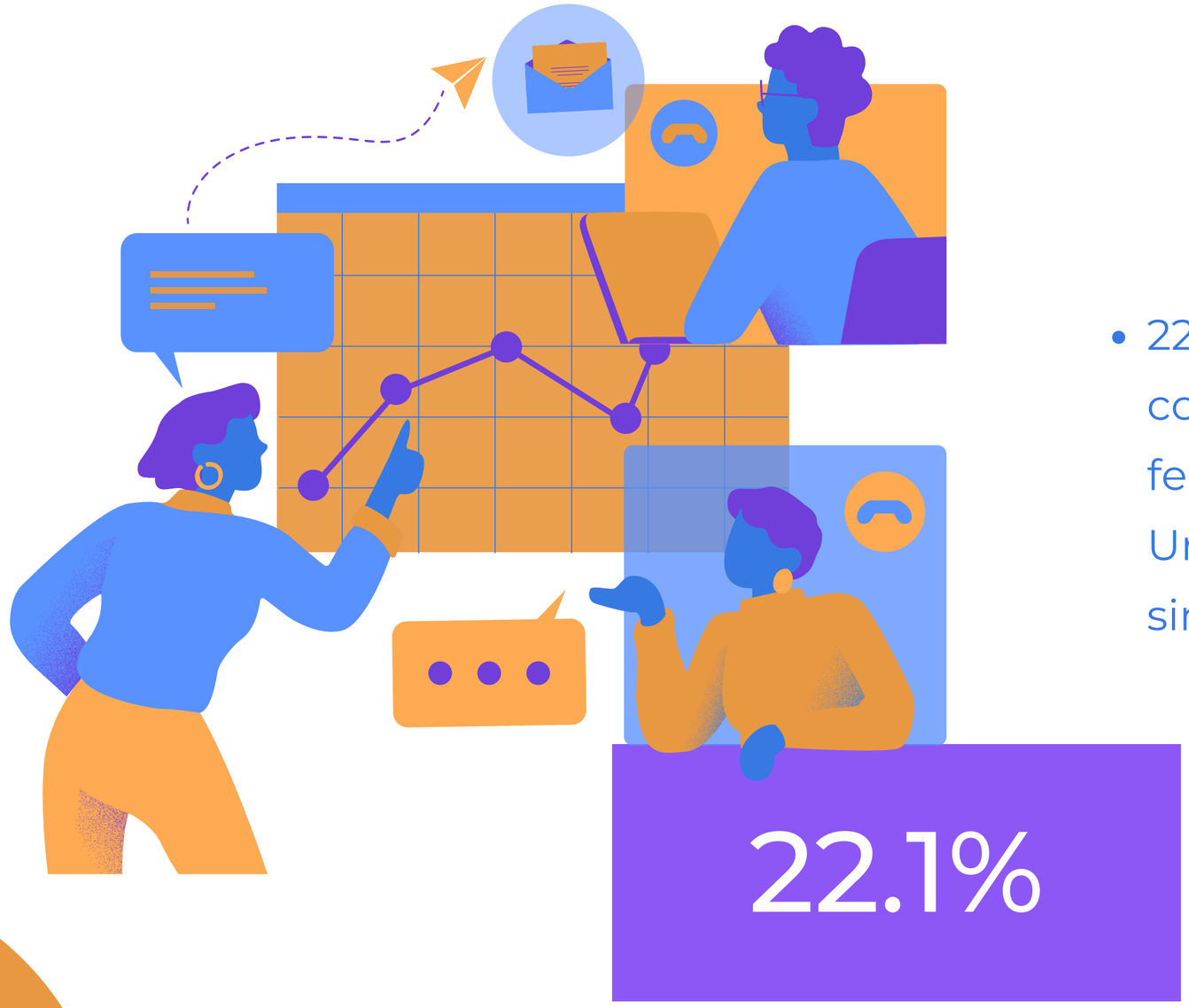




What is the overall distribution of key features?

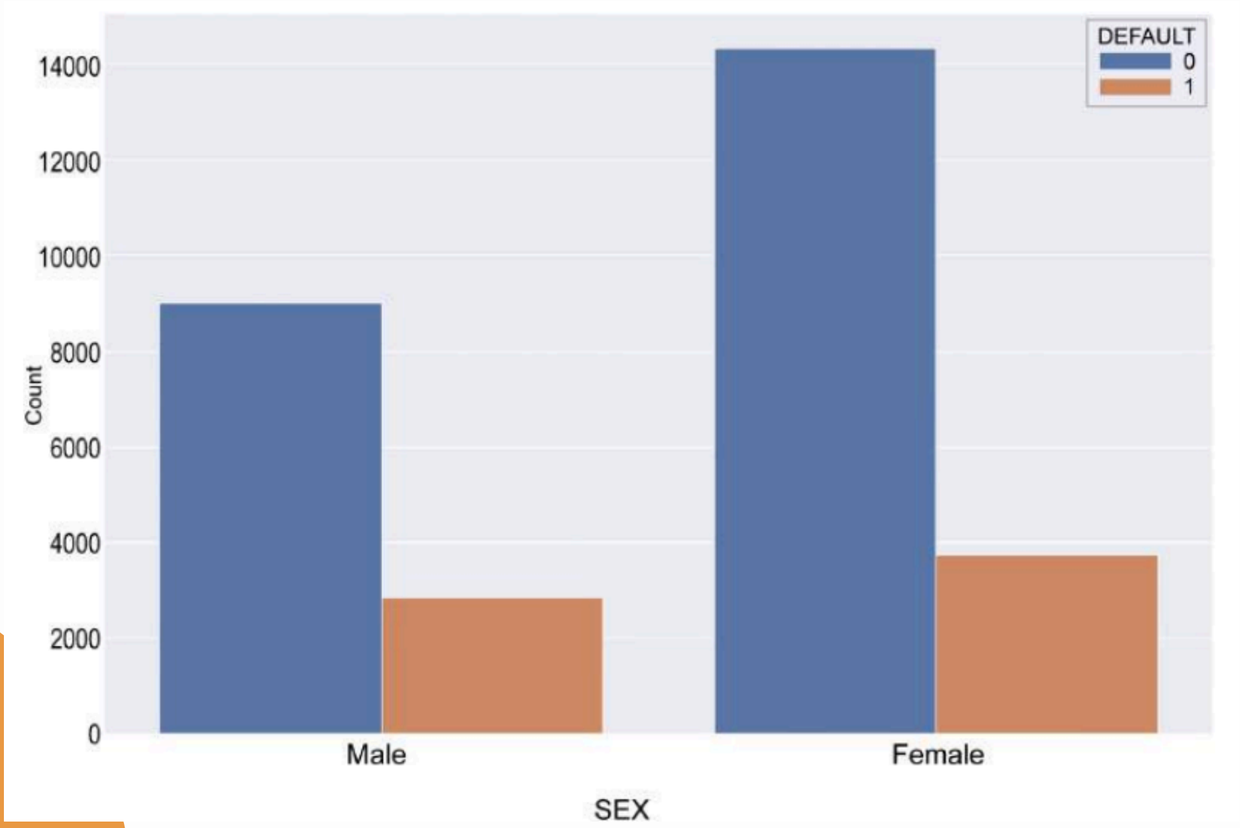






- 22.1% of customers default (DEFAULT countplot). Slightly more males than females, the largest education group is University, and most customers are single/married.

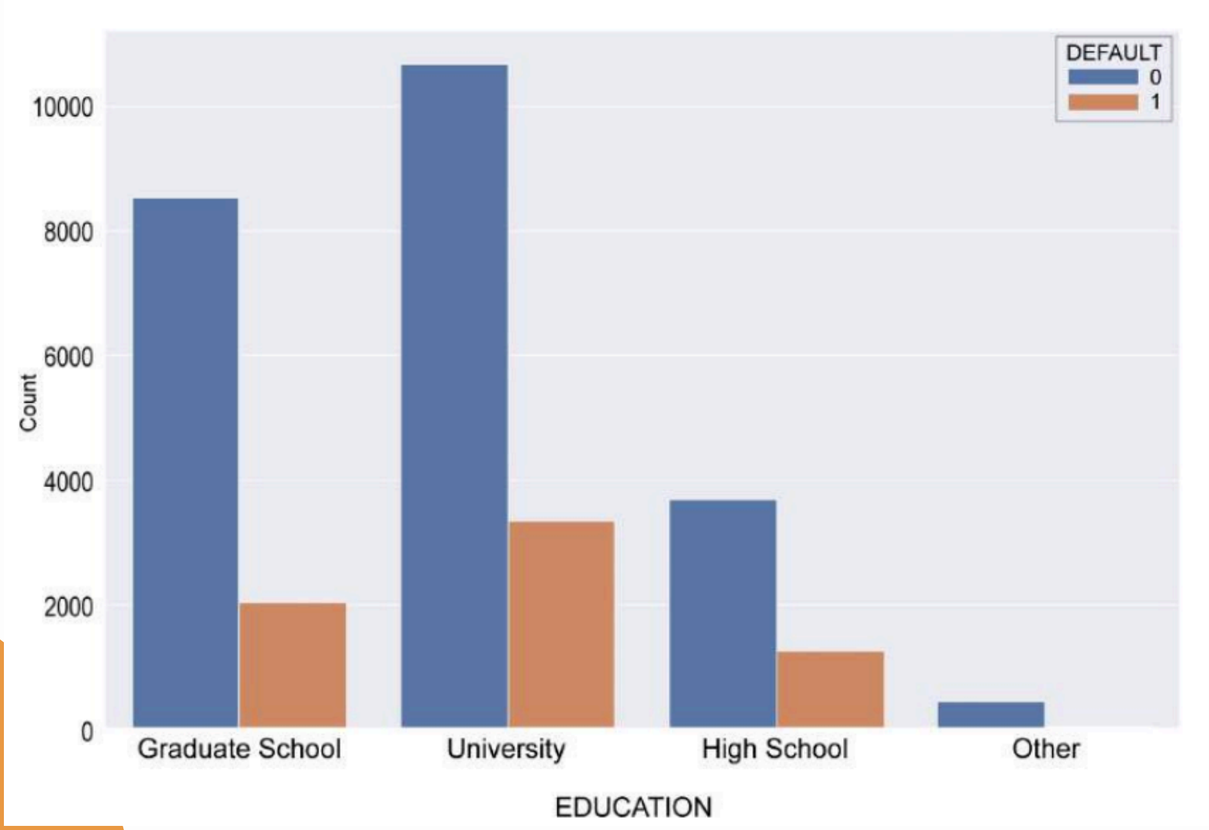
Do males default more than females?



- 24% of males default vs 20% of females. Males are at higher risk.

DEFAULT	0	1
SEX		
Female	0.792237	0.207763
Male	0.758328	0.241672
All	0.778800	0.221200

Does education level affect default?

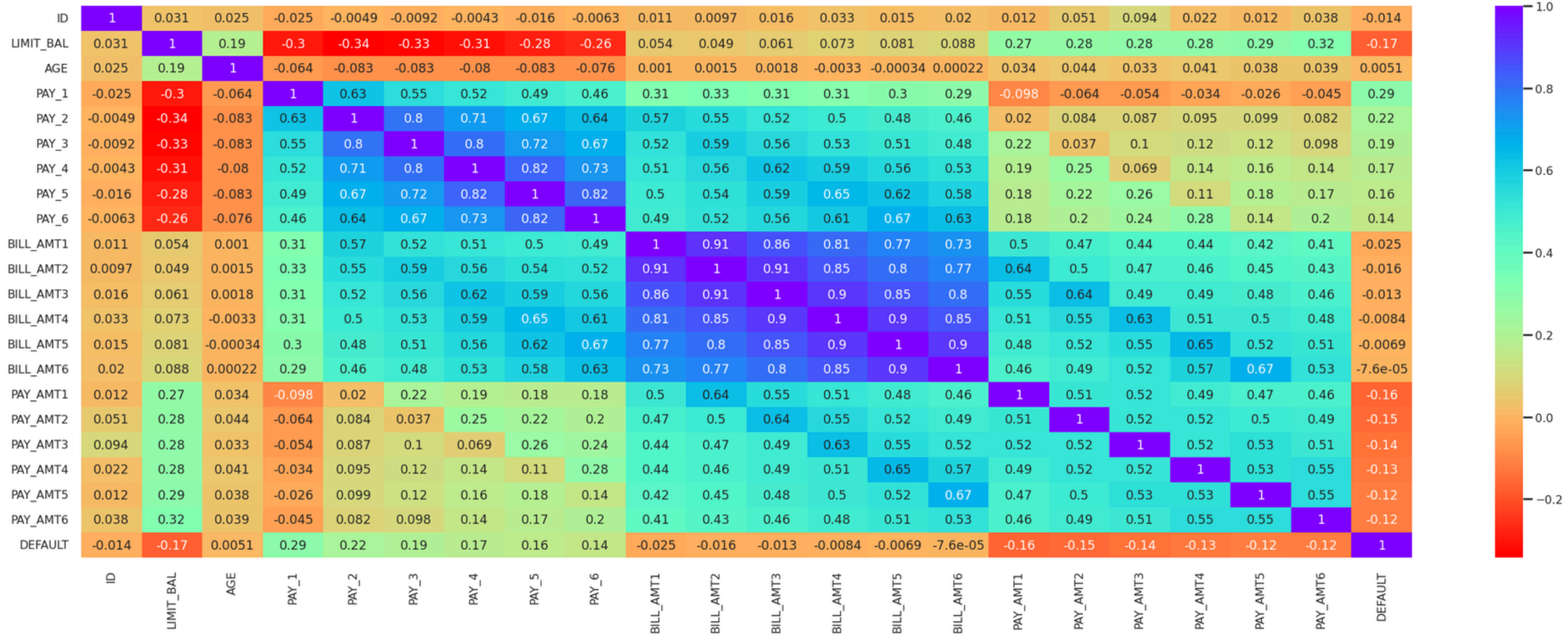


- High-school graduates default ~25%, higher than University (~20%) or Graduate (~18%).

DEFAULT	0	1
EDUCATION		
University	10700	3330
Graduate School	8549	2036
High School	3680	1237
Others	435	33

Correlation of Features with DEFAULT

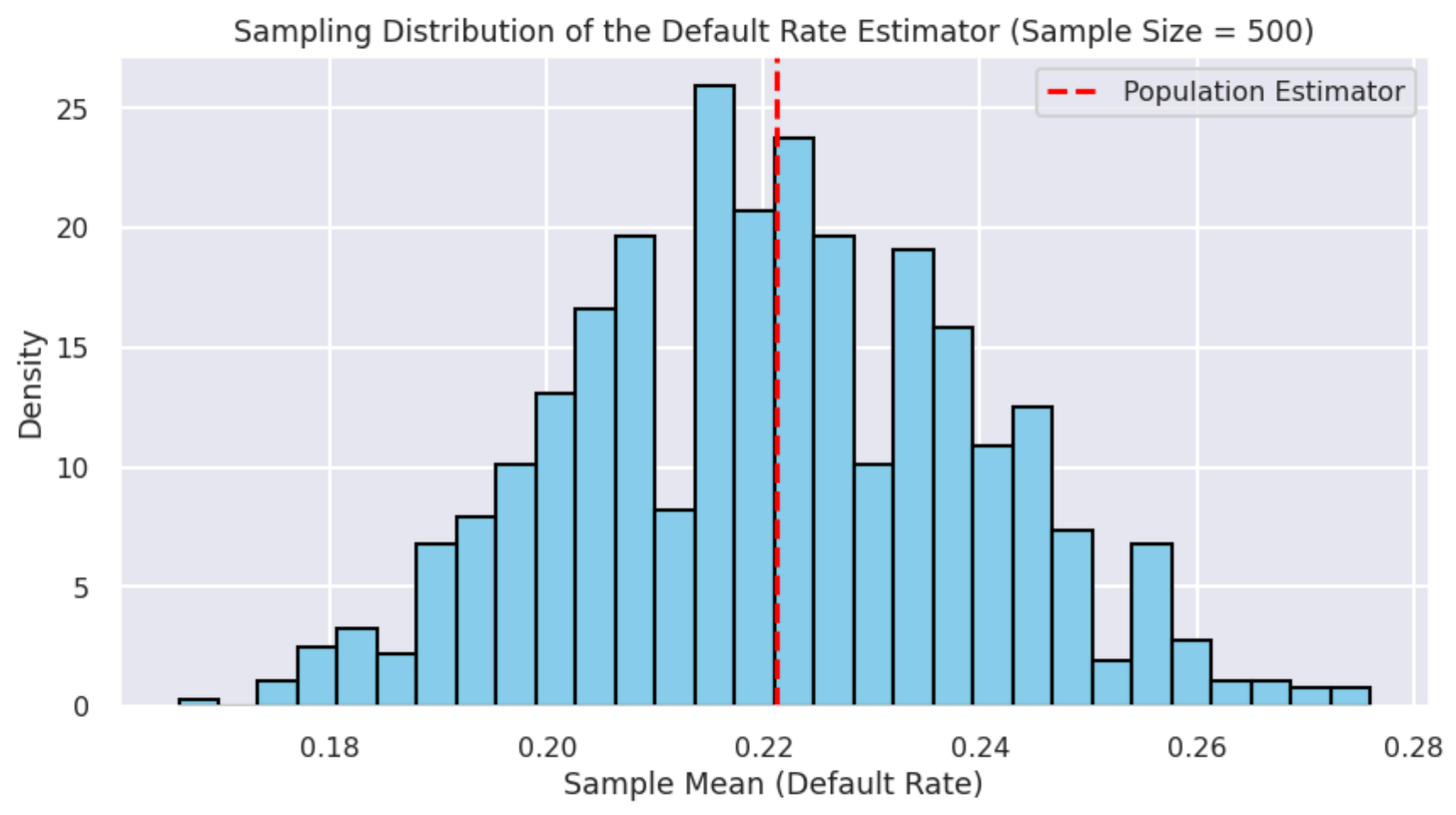
- We checked Spearman correlations. PAY_1 (0.29) and PAY_2 (0.22) are strongest predictors; PAYMENT AMT negatively correlated.



How reliable is our default-rate estimate?

Sampling Distribution of Default Rate

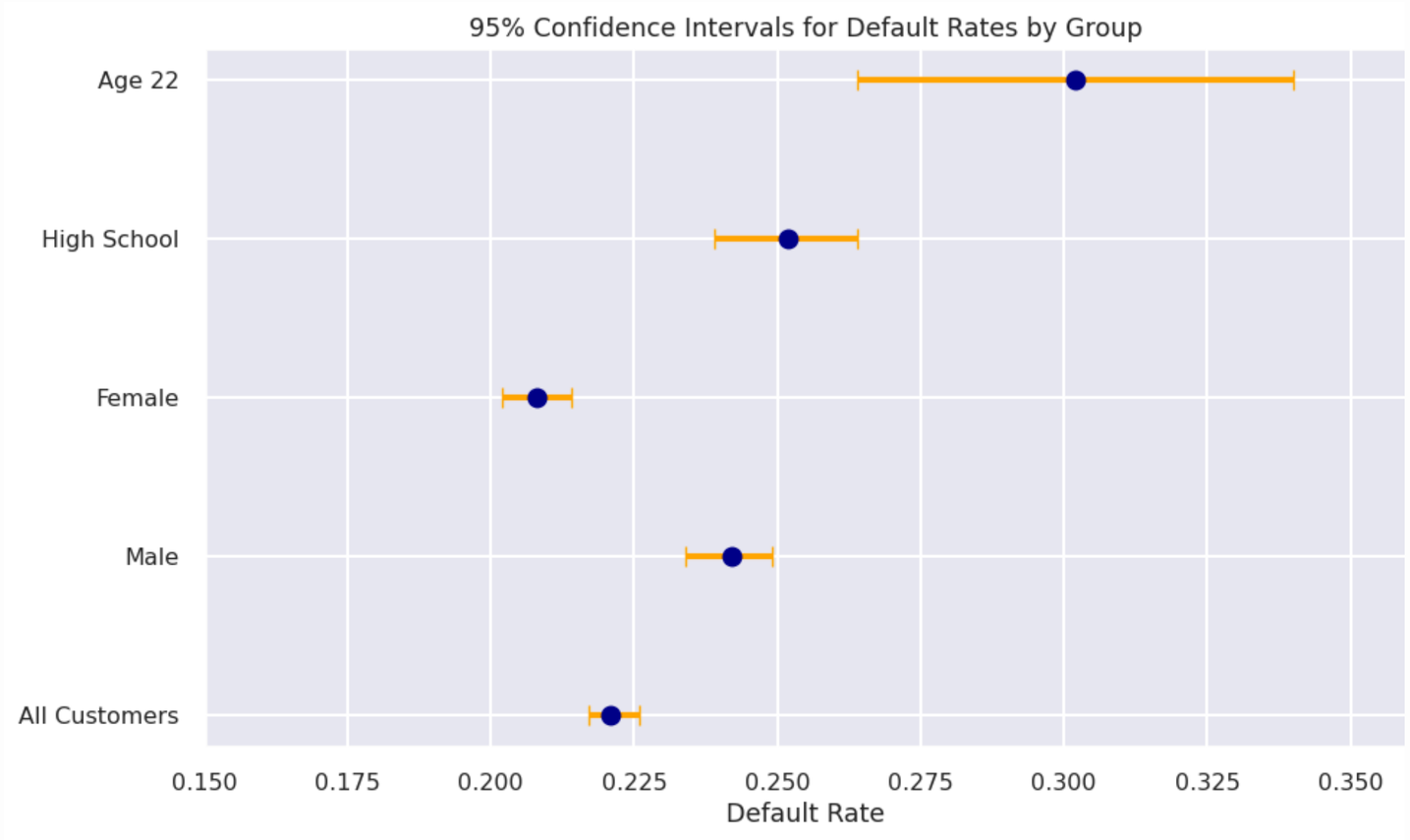
- We simulated 1,000 samples of 500 customers. The distribution of sample means is centered at ~22.2% with low variance, indicating our estimator is unbiased and stable.



WHAT IS THE UNCERTAINTY AROUND OUR ESTIMATES?

95% CIs for Default Rates

- All customers: 21.7%–22.6%. Males: 23.4–24.9%. Females: 20.2–21.4%. High School: 23.9–26.4%. Age 22: 26.4–34.0%.



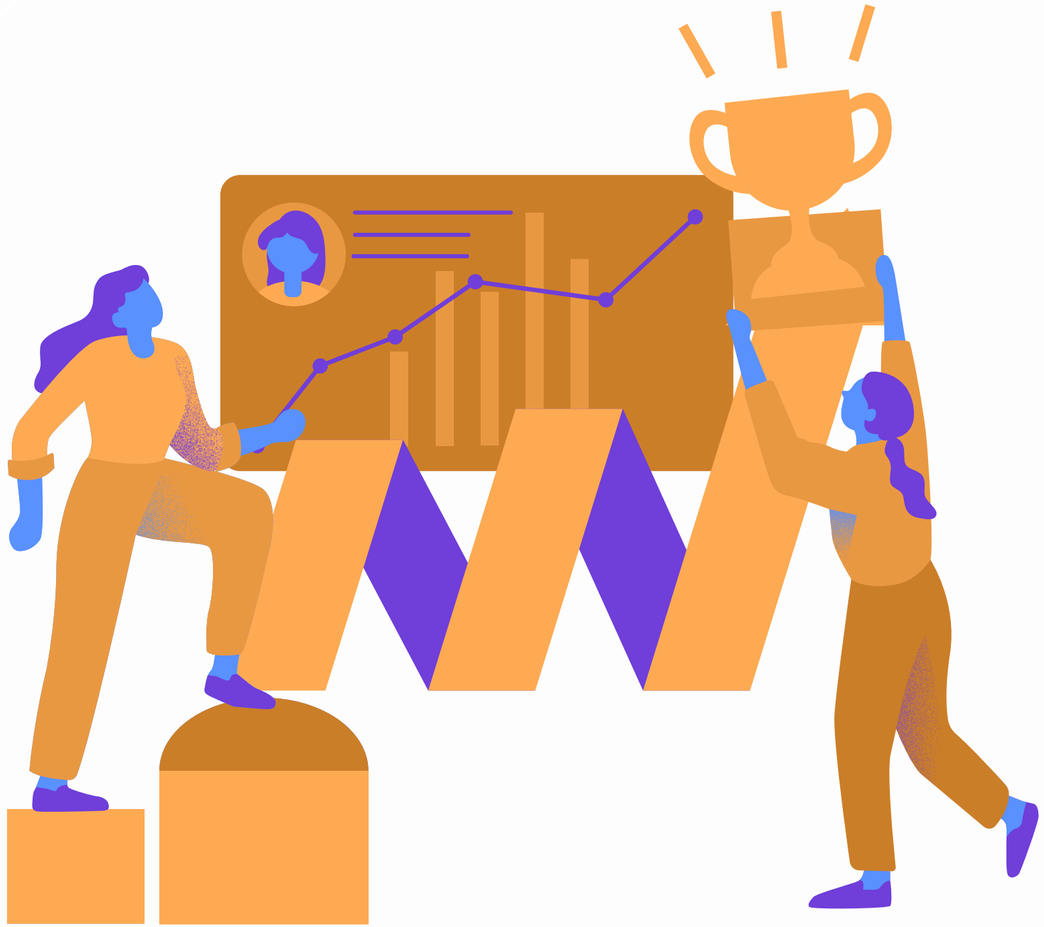
ARE OBSERVED DIFFERENCES SIGNIFICANT?

Z-Tests for Key Comparisons

- Males vs females: $Z=6.92$, $p<0.001$ → significant. HS vs Univ: $Z=2.01$, $p=0.045$ → borderline significant. Age 22 vs others: $Z=4.64$, $p<0.001$ → significant.

Conclusion & Actions

- High-risk customers: males, high-school educated, age 22, repeated late payers. Our combined EDA & inference workflow validates these insights.



Thank you.