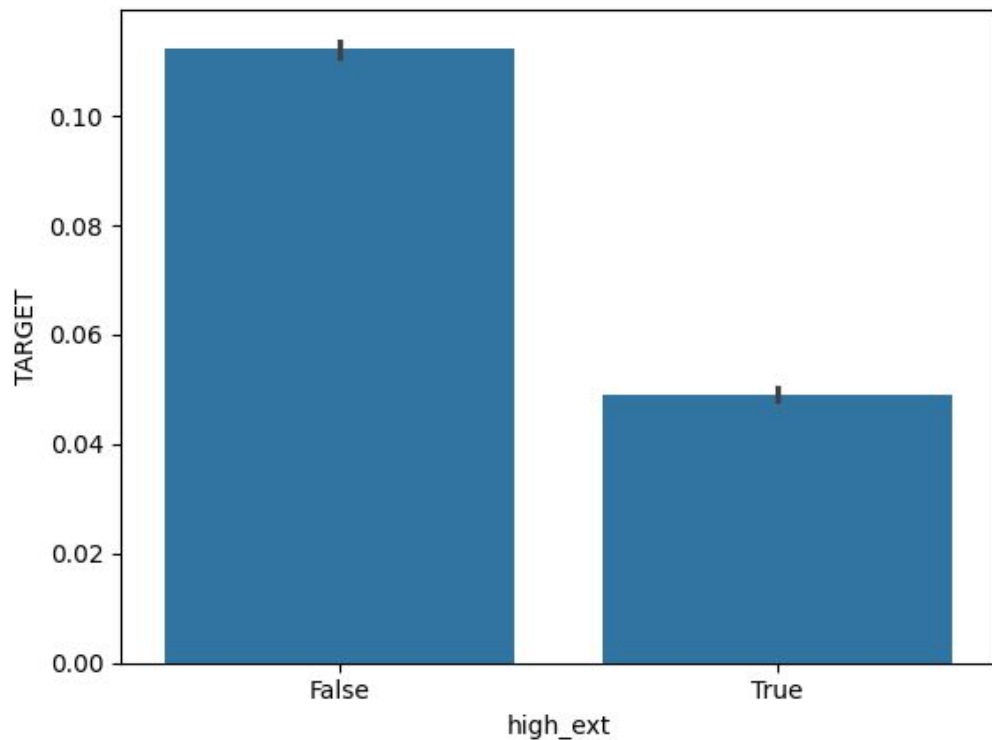# EDA Findings

Jasmine Zhang, Ishaan Banerjee, Zien Zhu

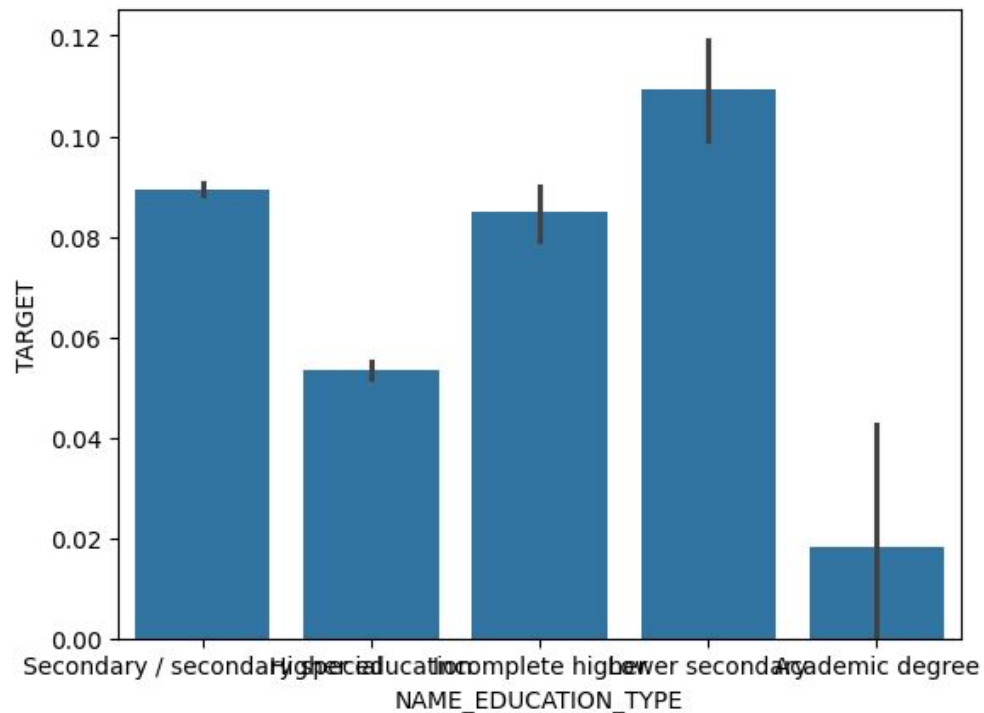# External Credit Score (EXT_SOURCE_2)

Strong relationship between the external credit score and default probability.

Applicants with below-median external credit scores have a default rate of 11.2%, compared to above-median credit scores of only 4.9%.

This is more than double increase in default risk, so external credit score the strongest predictor identified in my analysis.
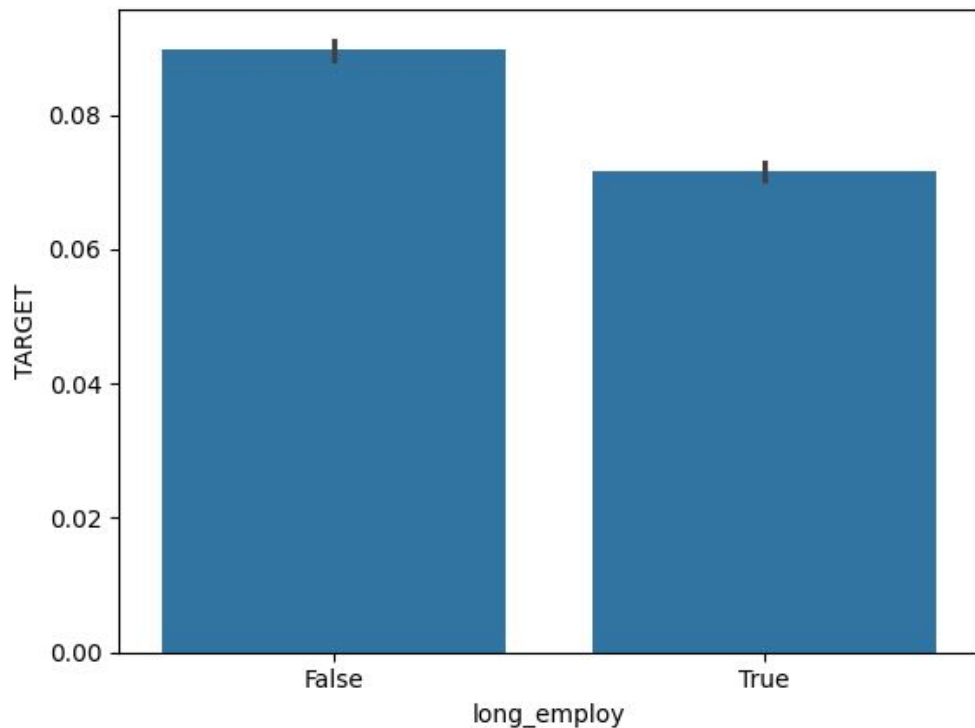
Education Level (NAME_EDUCATION_TYPE)

Default risk decreases steadily as education level increases. Lower secondary education has a 10.9% default rate, secondary education has a 8.9% default rate, higher education has a 5.4% default rate, academic degree has a 1.8% default rate.

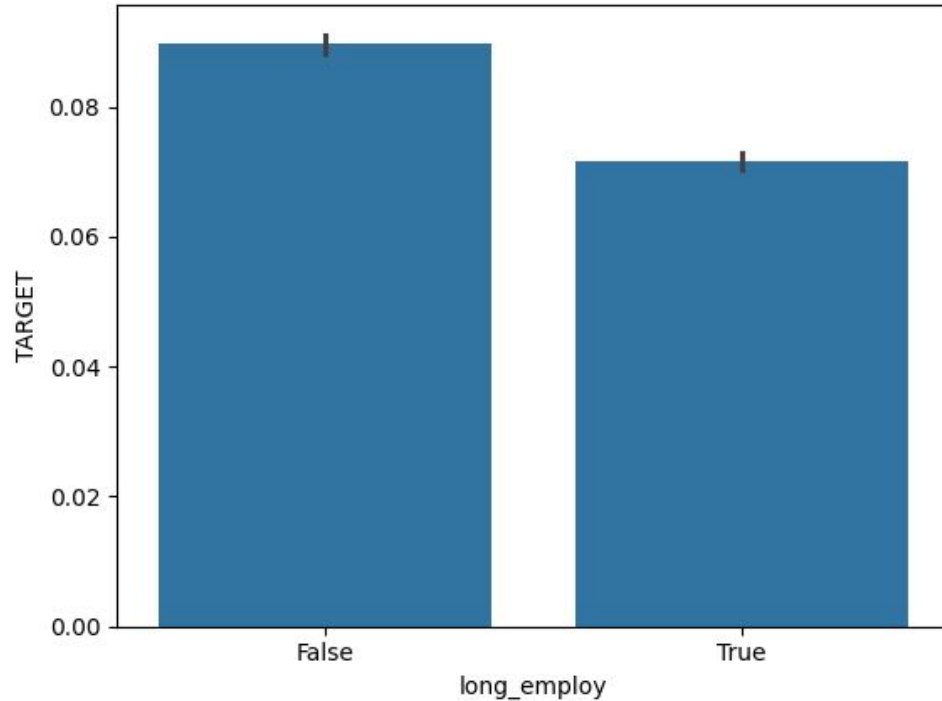Applicants with lower education have roughly double the default risk

Employment Length (DAYS_EMPLOYED)

Short employment had a 8.97% default rate and long employment had a 7.17% default rate.

Moderate increase of 15% default risk when going from long employment to short
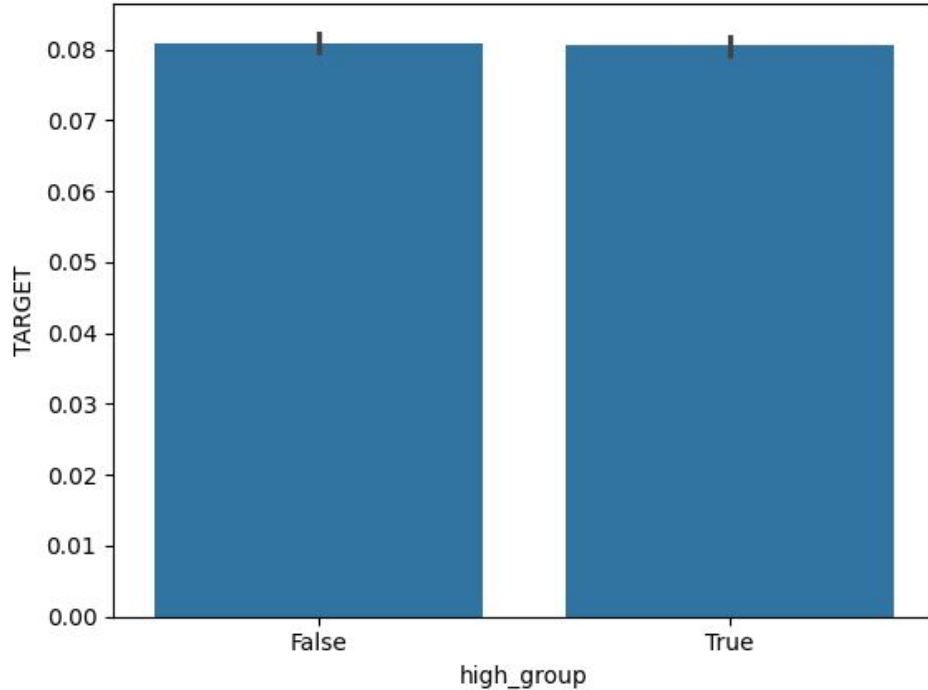
Income Type (NAME_INCOME_TYPE)

For large groups, Working hsa a 9.6% default rate, Commercial associate has a 7.5% default rate, State servant has a 5.8% default rate, Pensioner has a 5.4% default rate.

Extremely high default rate for Maternity and Unemployed, but so few data entries that just a few people defaulting artificially skews the data very high, so this is not statistically significant. These groups are too small to be reliable.
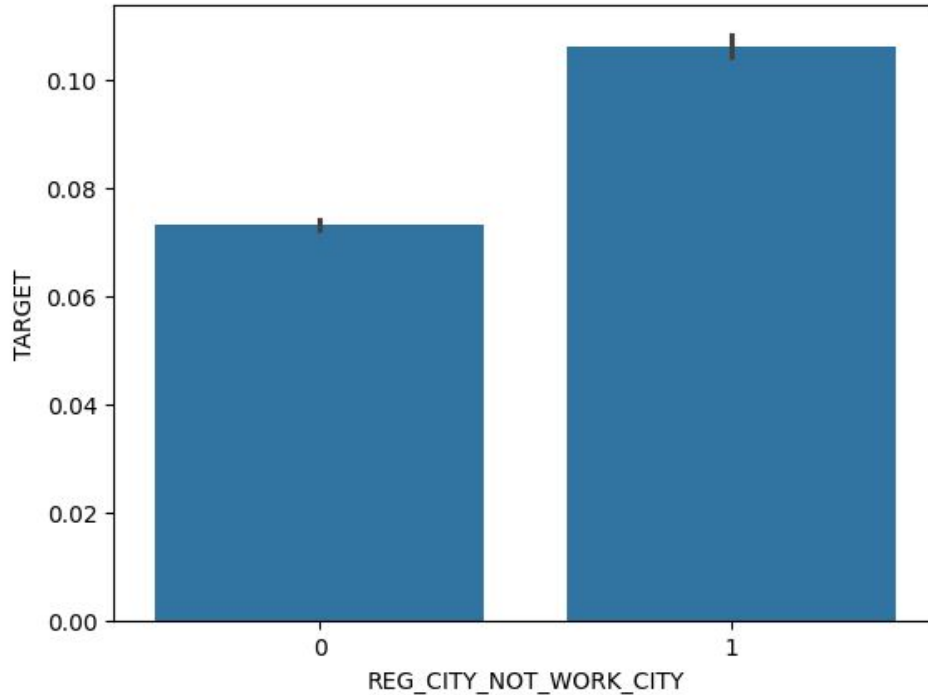
Loan Size (Loan-to-Income Ratio)

Borrowers split into two groups by median

Difference was negligible, suggests loan size relative to portion of borrower's income does not determine default risk.

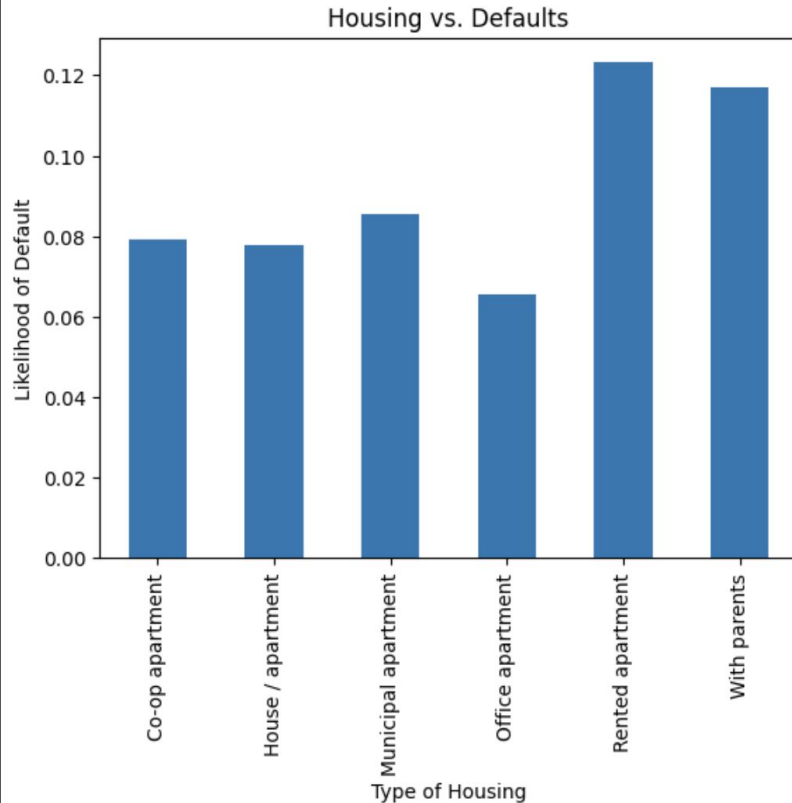# City Registration Mismatch (REG_CITY_NOT_WORK_CITY)

Borrowers whose registered city is different from their work city have slightly higher default rates.

Difference is moderate, registered city being different suggests some instability.

Default rate increases from 7.31% to 10.61% when registered city doesn't match work city, which is a 49% increase in default risk.

Housing Type (NAME_HOUSING_TYPE)

Housing vs. Defaults

Individuals in more stable or ownership-oriented housing, such as office or co-op apartments, maintain the lowest default rates, ranging between 6.5% and 8.5%.

A significant upward trend in risk appears for those in rented apartments or living with parents, where the likelihood of default peaks at over 12%.