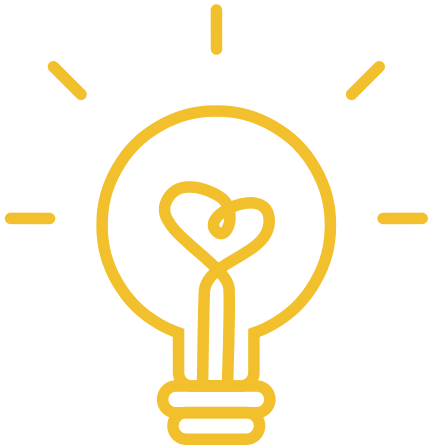
# LENDING CLUB

CASE STUDY

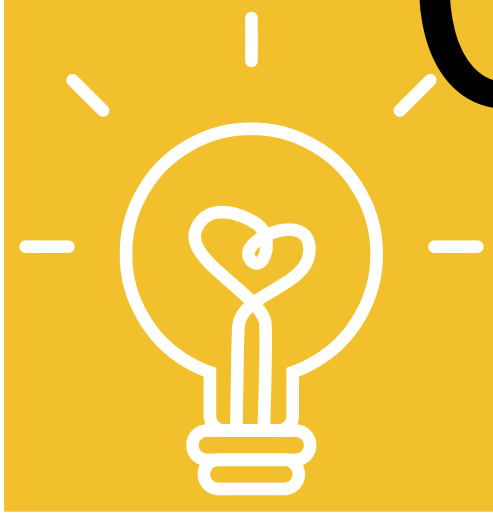
MAHMOUD NABEEL
UPGRAD AI/ML EPGP

The case study of Lending Club has a single goal in mind, which is identifying likely defaulting loans



# 01 Data Cleaning

It included removing columns with null/na/single-valued data
It also included deriving 3 new columns
Finally, I handled missing values



# 02 Handling Outliers

Put an assumption on specific columns for being outliers, and run some plots and quartiles removals to address them

# 03 Analyzing Data

Running univariate, segmented univariate and bivariate analyses on main data fields that I assume are major players in defaulting, which are interest rate, dti, purpose of loan, annual income and employee length

# CONCLUSIONS

## 01 MAIN REASONS OF DEFAULT

DTI has little impact
Higher interest rate, debt consolidation purpose, annual income between 31k-50k and lower employment length are the major risks to defaulting

## 02 OTHER REASONS

- People having annual income in the range 50000-100000.
- People having Public Recorded Bankruptcy.
- People with least grades like E,F,G which indicates high risk.
- People with very high Debt to Income value.
- People with working experience 10+ years.

## 03 Technology Used

I used the below python modules:
- Numpy
- Pandas
- MatPlotLib
- NBFormat
- Seaborn
- Plotly Express

## 04 Lessons Learned

All methods of analysis are important, and it helps navigate deeper into the dataset. Starting with univariate, followed by segmented univariate, then bivariate is a very logical order. Finishing up with a heat-map describes all relations in a nutshell.

# THANK YOU