**Slide 1**

Hello everyone, today I will discuss the implications of the provided credit risk dataset and my recommendations.

**Slide 2**

For an executive version of this presentation.   
My recommendation is that this dataset should not be used for decision-making and that the firm is better of recollecting data with proper predictors of credit risk instead. Additionally, the inclusion of the metadata would’ve been greatly appreciated.

First, the usage of certain demographic data is unethical and potentially illegal. The Consumer Financial Protection Bureau (CFPB) states that lenders cannot deny credit based on sex(gender), marital status, age (if over 62yo), or income source (s).

Second, as mentioned before, none of the metrics provided are strong predictors (or even directly related to) credit risks. The 5Cs of credit risk

1. **Capital**:
2. **Character:**
3. **Capacity:**
4. **Conditions:**
5. **Collateral:**

Third, static data does not tell the entire picture since purchasing behaviors change over the course of a year. For example, car sales often peaked during tax return season. Consequently, metrics such as monthly loan type distribution would be quite useful.

Nevertheless, here are the findings of the current dataset at hand. It is worth repeating that the implications of these findings are questionable at best and should not be generalized.

**Slide 3**

The logistic regression demonstrates that only four variables are significant in predicting a customer’s credit risk status, two of which are insignificant. Namely, (1) customers who apply for car loans (both used and new), (2) customers who rent, and (3) customer lifetime.

**Slide 4**

Viewing the relationship between customer lifetime and new/used car loan application. Results show that new customers prefer new cars whilst longtime customers prefer used cars. Unexpectedly, age was not a contributing factor to this relationship.

Grouping by credit risk, only the difference between low and high-risk customers applying for used cars was significant in that high-risk customers generally have higher lifetime than that low-risk customers.

**Slide 5**

Demographic trends show that divorcees, young adults, non-house owners, and long-term customers are at higher risk of defaulting on credit. Here are the potential explanations for these findings:

For divorcees:

* Victim of revenge as ex-spouse could drag credit scores down via large purchases on joint account(s).

For young adults, their young age makes them more vulnerable to high credit risk

* No/shorter credit history.
* High delinquency rates on credit card payments

Finally, as customers maintain longer-term relationships with the credit union or bank, the organization becomes more inclined to approve credit loans.

**Slide 6**

From these results, potential recommendations include.

1. Risk-based pricing: the lender could increase interest on high-risk demographics. However, due to regulation constraints, this approach might have limited use in predictive analyses.
2. Longer Loan Terms: Interestingly, the terms of all loan applications in this dataset ranged from 1 to 4 years which limited certain categories such as business, education, and housing loans. Additionally, short-term loans also have higher risks of default so lenders should consider extending loan terms for the above categories.
3. Credit insurance: lender could offer credit insurance. However, due to the high cost, options are extremely limited with this approach. Examples include clients who can’t defer/put into forbearance, or debt after death.
4. Diversifying borrower pool: one noticeable trend in demographic distribution is that all women are divorcees. Furthermore, there are big biases towards certain groups as well (e.g., homeowners). Given the case of Silicon Valley Bank going bankrupt partly due to its overreliance on startups. Lenders might consider diversifying their borrower pool.
5. Promotions: lender might consider demographic target promotions. For example, car loan promotions during tax return seasons.

**Slide 7**

3W:

**What went well:**

It was a lot of fun trying to understand this oddball of a dataset. I and Robert were discussing whether this data contained approved or pending applications and the former makes more sense. However, it would have meant that the lender has an agenda against non-divorced women.

The other thing was more fancy charts.

**What did not go well**

I did not have the time to create a dashboard

Writing the script made me realize that I could have focused more on understanding the results of the logistic regression, creating a smoother flow.

I finished this project with more questions than answers. An example is the aforementioned discrimination against non-divorced women.

**Improvement**

I am content with my results and I have come to accept the cruel reality of data analytics in the real world.