Loan Default Exploratory Analysis

Team: Rose Ro, Carol Yin, Kucina Li, Ziwen Ding

Date: Dec. 10th, 2020

Agenda

- Background
- Business Use Cases
- Data processing Tools
- Database Design
- Borrower Risk Profiling with Tableau
- 5C’s Credit Analysis Approach with SQL & Tableau

Background

- Loans are created and renewed between borrowers and lenders monthly. These loans have different conditions based on different attributes of the borrowers and behave differently as time passes.
- **Companies spend time understanding borrower behavior and the default probabilities to minimize risk and save money.**

- **Our Goals:**
  a. Build highly normalized (3NF+) data warehouse and EER diagram **so that the loan company can extract, store and analyse data more efficiently**
  b. Use the large loan dataset, 20,000+ records incurring each quarter, to **analyse the demographic of borrowers, the likelihood of default, and critical features related to the default.**

Business Use Case

As a loan company, we plan to explore implicit relationships between different types of data points in a complete set of loan data to:

Risk Management

Determine whether a loan will default and predict the loss incurred for the investors if it does default.

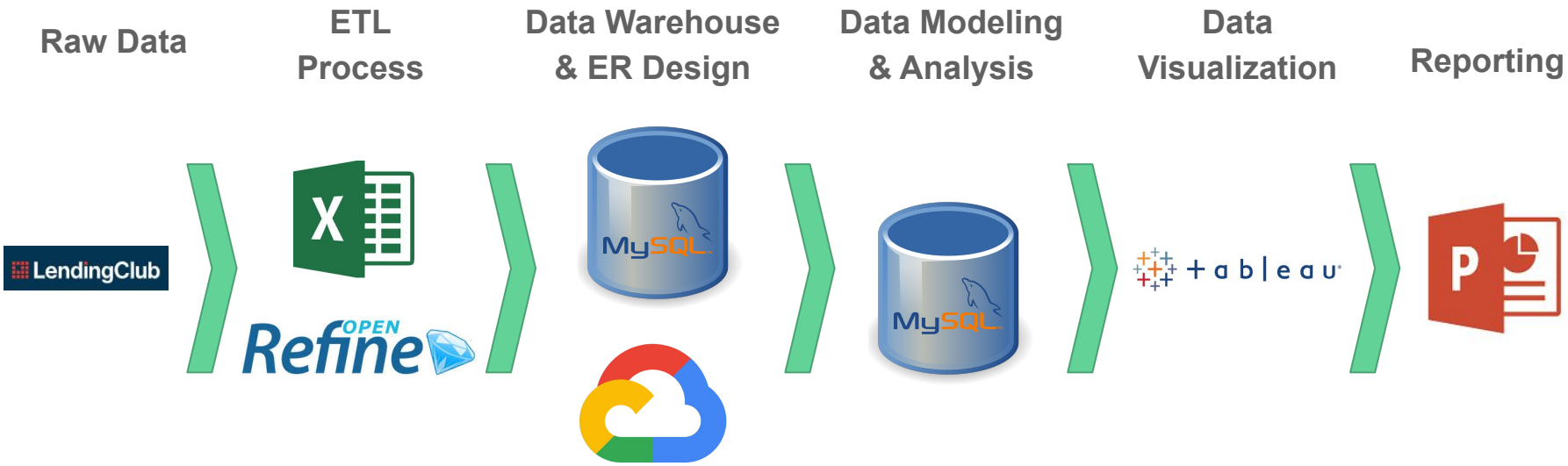
ROI Optimization

Offer tailored loan plans for customers with different risk types in order to maximize investors profit .

Data Source

Data Set: Lending Club
Time Period: 2020 Q2
Details: loan info, loan payment account info, borrower info…
Dimension: 150 Columns, 13000+ Rows

Data Processing Tools



Database Design - Platform Considerations

- OLTP database
- Data Consistency and Accuracy
- Provided indexed data sets for rapid searching, retrieval, and querying
- Supported the large dataset by backup and recovery, as part of a high availability strategy, can be performed on a low level of granularity to efficiently manage the size of the database.
- Ensured relational data integrity by correctly placing PK, FK constraints and NOT NULL definitions

Database Design - Data Preparation Steps

- Initial data cleaning: reduced from 150 columns to 93 columns by deleting blank, mostly 0 and unimportant columns.
- Deleted attributes which represent redundant information to have an efficient structure.
- Examined the cardinality between the entities.
- Turned character variables to binary variables for easy calculation, for example, “Y/N” to “1/0”.
- Identified three separate sub tables and converged them into one table with an additional attribute as sub-type.

Database Design - Attribute Selection Challenges

last_credit_pull_d	The most recent month LC pulled credit for this loan		loan	
last_fico_range_high	The upper boundary range the borrower's last FICO pulled belongs to.		FICO	
last_fico_range_low	The lower boundary range the borrower's last FICO pulled belongs to.		FICO	
application_type	Indicates whether the loan is an individual application or a joint application with two co-borrowers	Category: individual/joint	application_type	
annual_inc_joint	The combined self-reported annual income provided by the co-borrowers during registration		joint	
dti_joint	A ratio calculated using the co-borrowers' total monthly payments on the total debt obligations, excluding mortgages and the requested LC loan, divided by the co-borrowers' combined self-reported monthly income		joint	
tot_coll_amt	Total collection amounts ever owed	What is a collection account? If you fall behind on payments, the lender or creditor may t	account	credit?
tot_cur_bal	Total current balance of all accounts		account	
open_acc_6m	Number of open trades in last 6 months		account	
open_act_il	Number of currently active installment trades		installment_trade	
open_il_12m	Number of installment accounts opened in past 12 months		installment_trade	
open_il_24m	Number of installment accounts opened in past 24 months		installment_trade	
mths_since_rcnt_il	Months since most recent installment accounts opened		installment_trade	
total_bal_il	Total current balance of all installment accounts		installment_trade	
il_util	Ratio of total current balance to high credit/credit limit on all install acct		installment_trade	
open_rv_12m	Number of revolving trades opened in past 12 months	A revolving trade account is an account that provides you with a credit limit you're allowe	revolving_trade	
open_rv_24m	Number of revolving trades opened in past 24 months	revolving	revolving_trade	
max_bal_bc	Maximum current balance owed on all revolving accounts		revolving	
all_util	Balance to credit limit on all trades		account	
total_rev_hi_lim	Total revolving high credit/credit limit		revolving	
inq_fi	Number of personal finance inquiries		account	
total_cu_tl	Number of finance trades		account	
inq_last_12m	Number of credit inquiries in past 12 months		account	
acc_open_past_24mths	Number of trades opened in past 24 months.		account	
avg_cur_bal	Average current balance of all accounts		account	
bc_open_to_buy	Total open to buy on revolving bankcards.		bankcard	
bc_util	Ratio of total current balance to high credit/credit limit for all bankcard accounts.		bankcard	
chargeoff_within_12_mths	Number of charge-offs within 12 months	Charge off: A charge-off is a debt, for example on a credit card, that is deemed unlikely to be collected by the creditor because the borrower has become	account	credit
mo_sin_old_il_acct	Months since oldest bank installment account opened		installment	
mo_sin_old_rev_tl_op	Months since oldest revolving account opened		revolving	
mo_sin_rcnt_rev_tl_op	Months since most recent revolving account opened		revolving	
mo_sin_rcnt_tl	Months since most recent account opened		account	
mort_acc	Number of mortgage accounts.		mortgage	
mths_since_recent_bc	Months since most recent bankcard account opened.		bankcard	
mths_since_recent_inq	Months since most recent inquiry.		account	
num_actv_bc_tl	Number of currently active bankcard accounts		bankcard	
num_actv_rev_tl	Number of currently active revolving trades		revolving	
num_bc_sats	Number of satisfactory bankcard accounts		bankcard	
num_bc_tl	Number of bankcard accounts		bankcard	
num_il_tl	Number of installment accounts		installment	
num_op_rev_tl	Number of open revolving accounts		revolving	
num_rev_accts	Number of revolving accounts		revolving	
num_rev_tl_bal_gt_0	Number of revolving trades with balance >0		revolving	
num_sats	Number of satisfactory accounts	accounts have been paid in full and on time	account	
num_tl_op_past_12m	Number of accounts opened in past 12 months		account	
pct_tl_nvr_dlq	Percent of trades never delinquent		account	credit?
percent_bc_gt_75	Percentage of all bankcard accounts > 75% of limit.		bankcard	
pub_rec_bankruptcies	Number of public record bankruptcies		account	credit?
tot_hi_cred_lim	Total high credit/credit limit		account	
total_bal_ex_mort	Total credit balance excluding mortgage		account	
total_bc_limit	Total bankcard high credit/credit limit		bankcard	
total_il_high_credit_limit	Total installment high credit/credit limit		installment	
hardship_flag	Flags whether or not the borrower is on a hardship plan		borrower	
total_bal_bc	total_bc_limit - bc_open_to_buy	new added	bankcard	

Used Google Sheets

Challenge 1:

Understanding financial terms (by using the data dictionary provided by lending club)

Challenge 2:

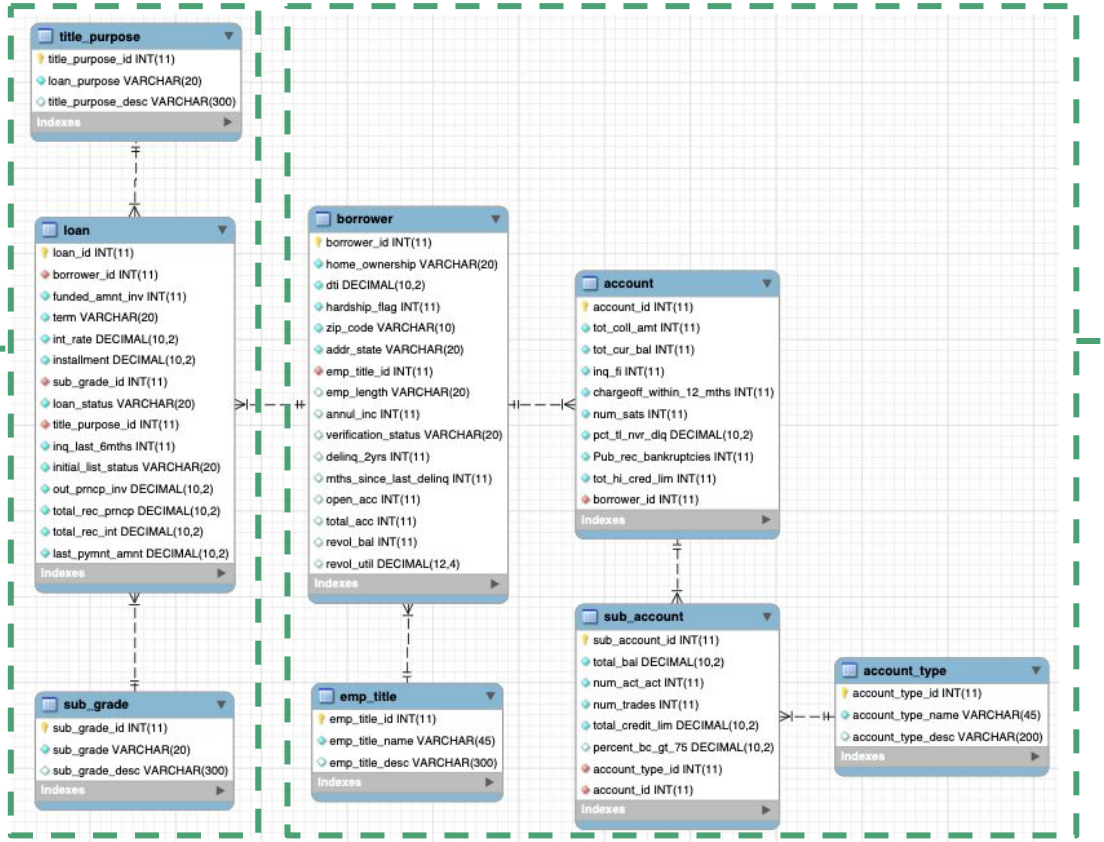
Grouping them to different entities

Challenge 3:

Deleting unimportant attributes

Database Design - EER Model

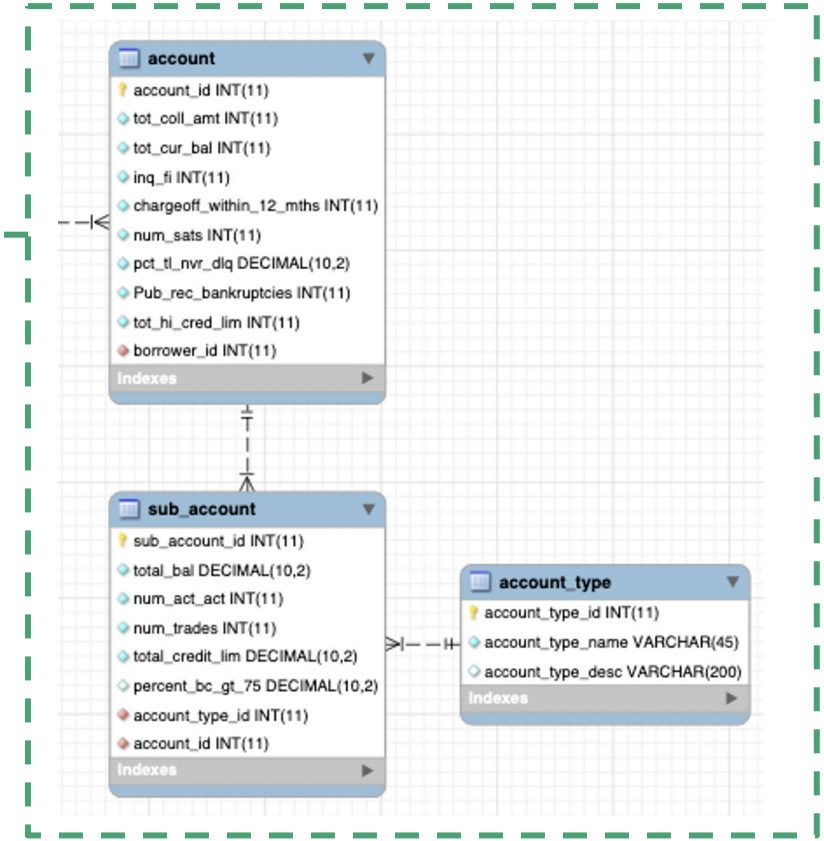
Loan



Borrower

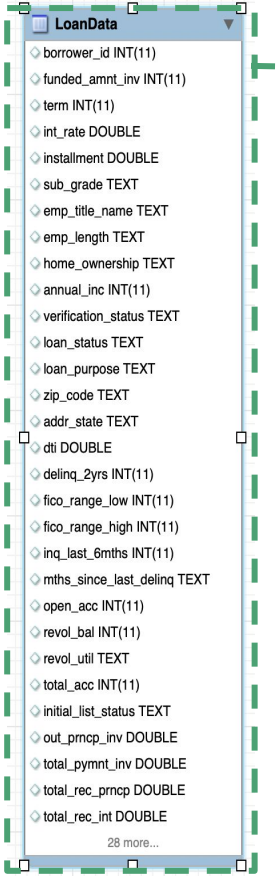
Database Design - EER Model

Borrower




```
• select
  A.borrower_id,
  1 as account_type_id,
  # total_bal_bc (column manually created)
  total_bc_limit - bc_open_to_buy as total_bal,
  num_actv_bc_tl as num_act_act,
  num_bc_sats as num_trades,
  total_bc_limit as total_credit_lim
  FROM
  staging.LoanData AS A
      LEFT JOIN
  staging.bc_open_to_buy AS B
      ON A.borrower_id = B.borrower_id

  union all

  select
  borrower_id,
  2 as account_type_id,
  total_bal_il as total_bal,
  open_act_il as num_act_act,
  num_il_tl as num_trades,
  total_il_high_credit_limit as total_credit_lim
  FROM staging.LoanData

  union all

   select
   borrower_id,
   3 as account_type_id,
   max_bal_bc as total_bal,
   num_actv_rev_tl as num_act_act,
   num_op_rev_tl as num_trades,
   total_rev_hi_lim as total_credit_lim
   FROM staging.LoanData;
```

Database Design - Insertion

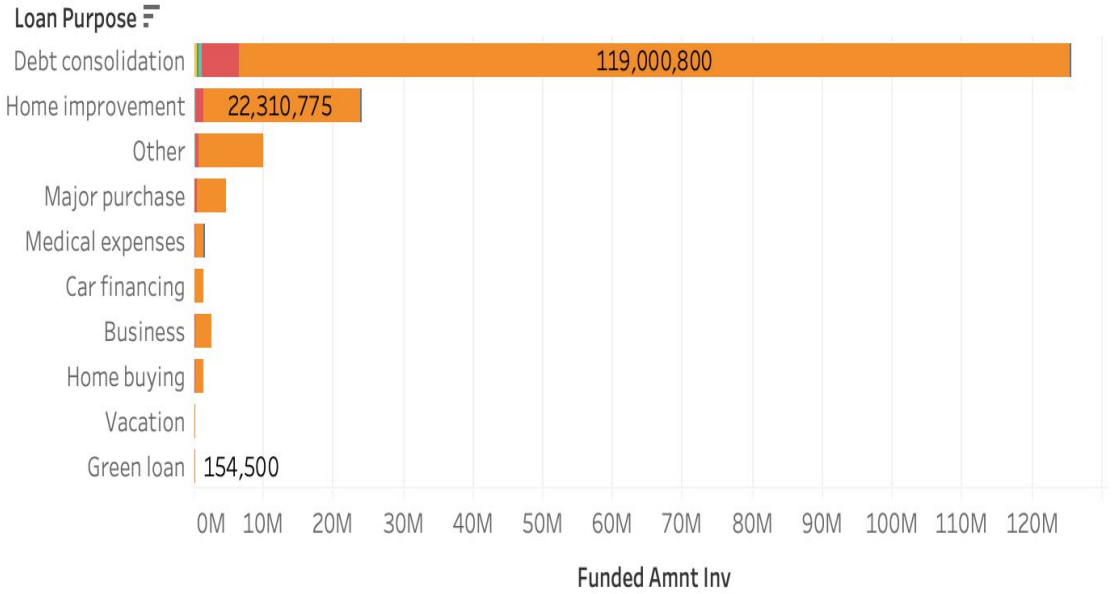


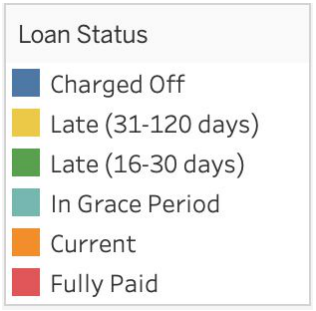
Staging Table

```
INSERT INTO default_rate_predict.borrower(borrower_id, home_ownership, dti, hardship_flag, zip_code, addr_state,
emp_title_id, emp_length, annul_inc, verification_status, delinq_2yrs, mths_since_last_delinq, open_acc, total_acc,
revol_bal, revol_util)
(SELECT L.borrower_id, L.home_ownership, L.dti, case when L.hardship_flag = 'N' then 0 else 1 end, L.zip_code, L.addr_state,
e.emp_title_id, L.emp_length, L.annual_inc, L.verification_status, L.delinq_2yrs, L.mths_since_last_delinq,
L.open_acc, L.total_acc, L.revol_bal, L.revol_util FROM staging.LoanData L
    INNER JOIN
default_rate_predict.emp_title e on e.emp_title_name = L.emp_title_name);

INSERT INTO default_rate_predict.account(tot_coll_amt, tot_cur_bal, inq_fi, chargeoff_within_12_mths, num_sats,
pct_tl_nvr_dlq, Pub_rec_bankruptcies, tot_hi_cred_lim, borrower_id)
(SELECT L.tot_coll_amt, L.tot_cur_bal, L.inq_fi, L.chargeoff_within_12_mths, L.num_sats,
L.pct_tl_nvr_dlq, L.pub_rec_bankruptcies, L.total_il_high_credit_limit, L.borrower_id FROM staging.LoanData L);
```

Insertion Code

Borrower Risk Profiling: Funded Amount by Purpose

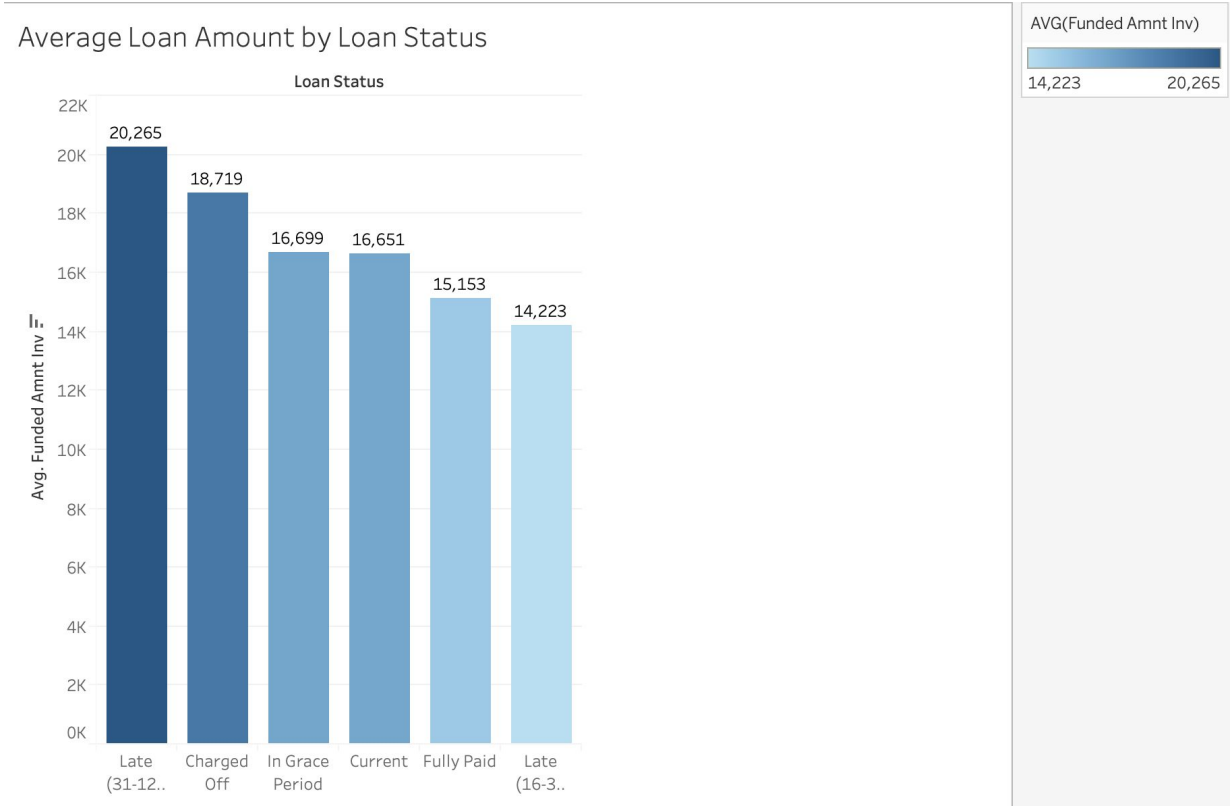




Conclusion: **Borrowers who take on loans for *debt consolidation purpose** heavily dominate the loan market.

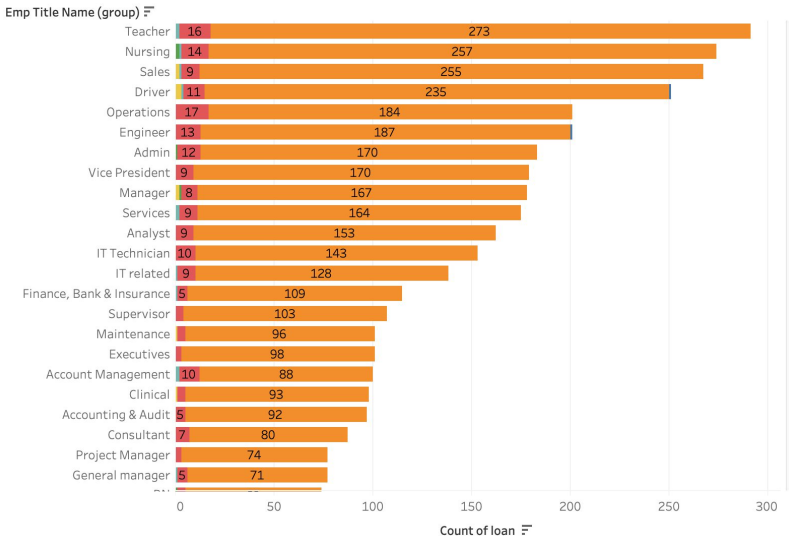
***Debt consolidation** allows you to combine multiple **debts** into a single balance with a single monthly payment

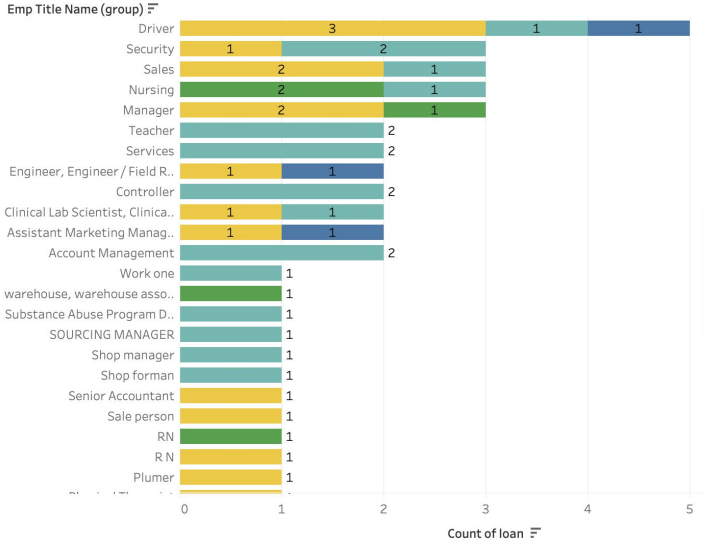
Borrower Risk Profiling: Average Funded Amount

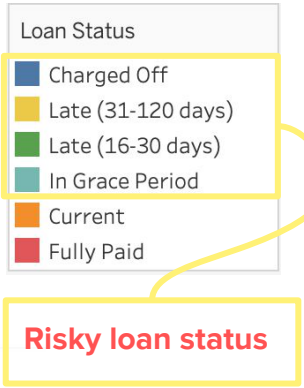


Conclusion:
Risky loan status (Late, Charged Off, In Grace Period) have higher average loan amount.

Borrower Risk Profiling: Count of Loans by Employment Type



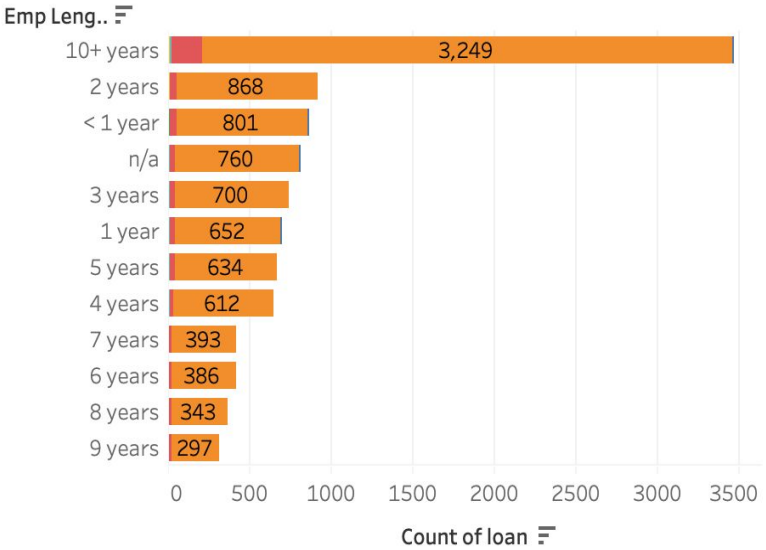


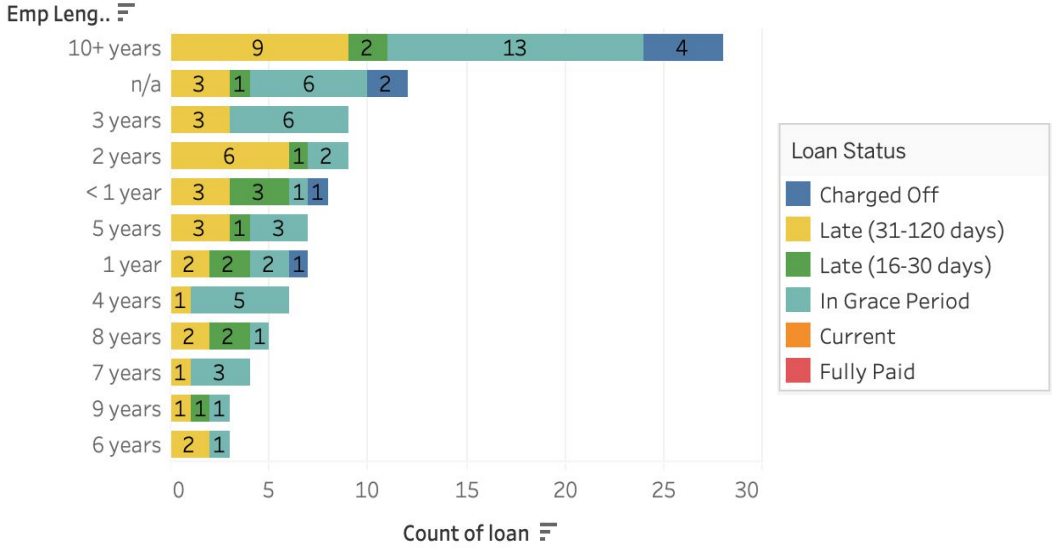


Overall, **teacher, nursing, sales, and driver** employment types borrow more loans

Driver, security, sales, and nursing employment types have more loans with risky loan status.

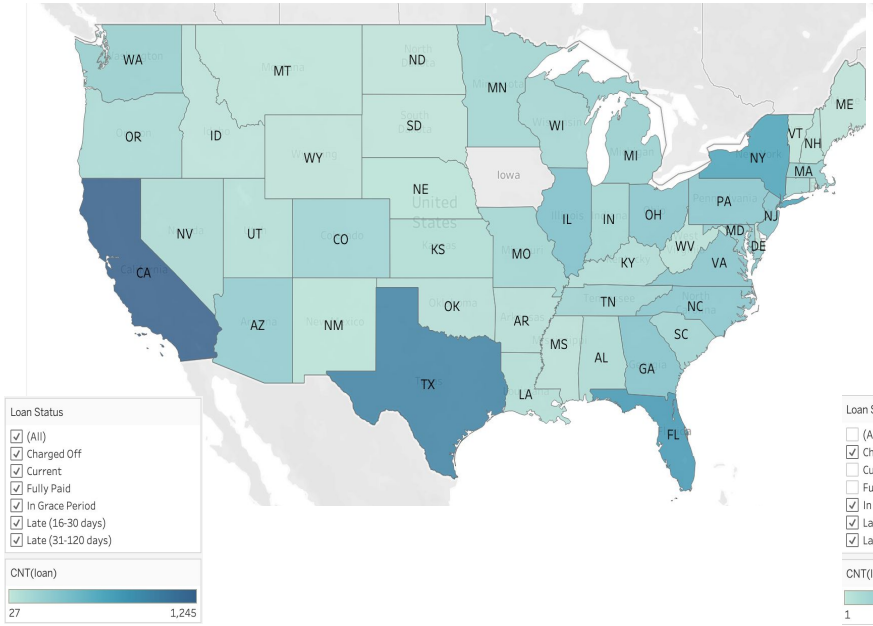
Borrower Risk Profiling: Count of Loans by Employment Length

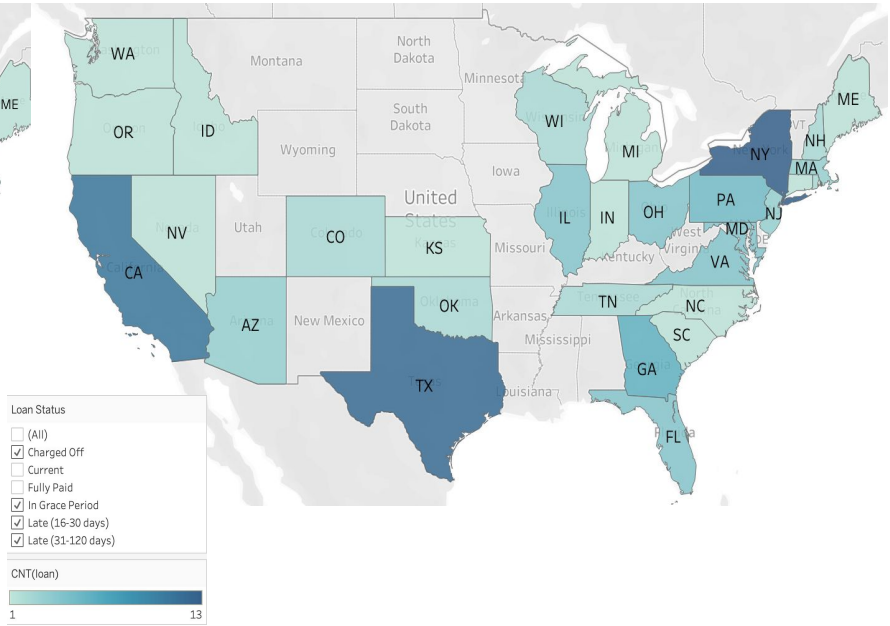




Conclusion: Whether it is all status comparison or with only the risky loan status,
borrowers with 10+ years work experience dominates the loan market.

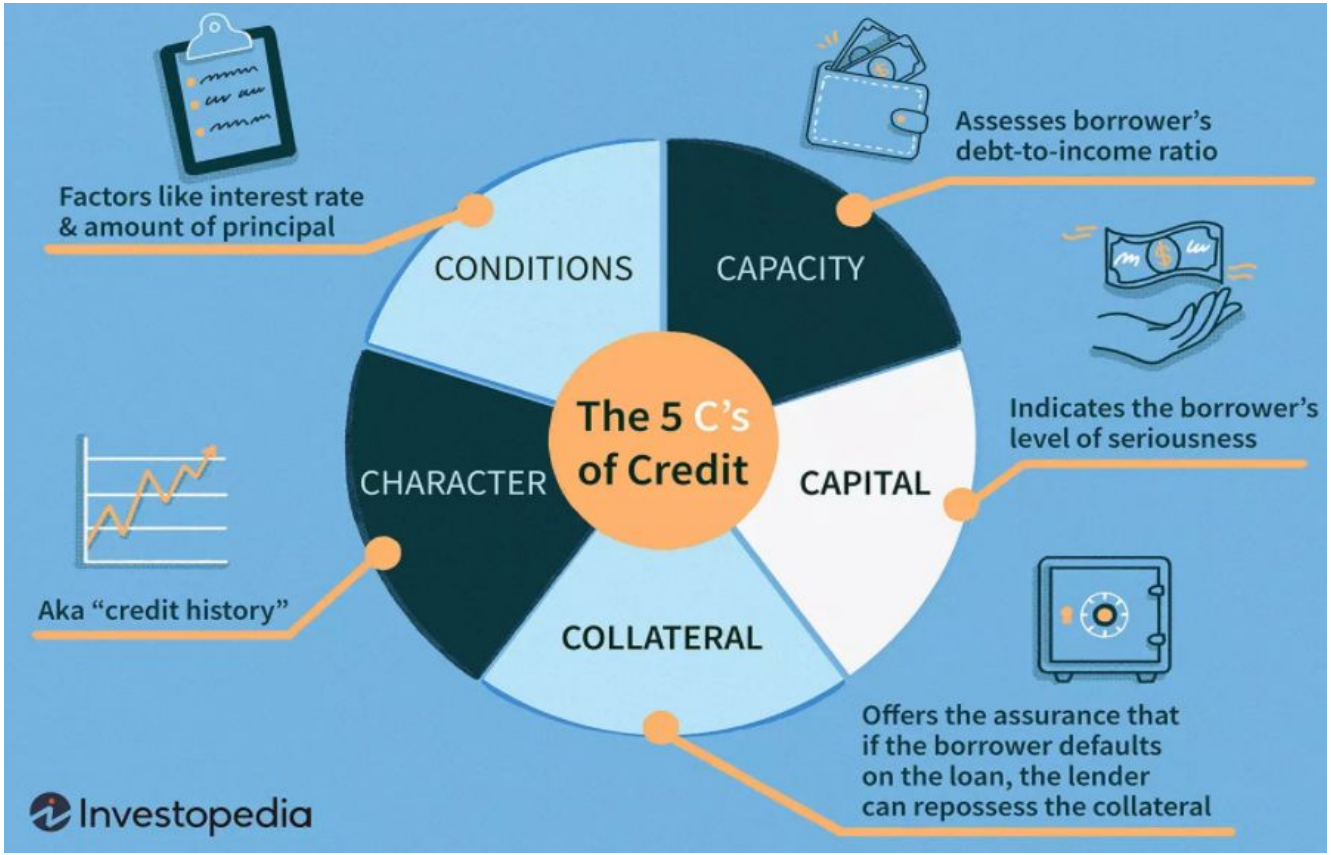
Borrower Risk Profiling: Count of Loans by State





Conclusion: **More developed states such as CA, TX, NY generate more loans**, TX and NY also have higher portions of loans with risky loan status.

5C’s credit analysis approach



5C’s credit analysis approach:

Find average account balance for borrowers who get loan with different sub grade levels.

A1	214521.0315
A2	193337.2399
A3	165404.7922
A4	208680.5876
A5	180619.5480
B1	194137.3442
B2	188545.7060
B3	177859.3291
B4	195458.8463
B5	176302.3775
C1	182387.5175
C2	176993.0140
C3	172321.3230
C4	182270.5440
C5	179207.0239

Customers with which kinds of home ownership are more likely to fully pay their loan on time?

own	paid_percentage
MORTGAGE	5.6853%
OWN	5.0405%
RENT	3.8275%

5C’s credit analysis approach:


```sql
SELECT
    S.sub_grade, AVG(A.tot_cur_bal) AS avgBalance
FROM
    account A
        INNER JOIN
    borrower B ON B.borrower_id = A.borrower_id
        INNER JOIN
    loan L ON L.borrower_id = B.borrower_id
        INNER JOIN
    sub_grade S ON S.sub_grade_id = L.sub_grade_id
GROUP BY S.sub_grade;
```

```sql
SELECT
    B.home_ownership AS own, CONCAT(m.count/COUNT(L.loan_id) * 100, '%') AS paid_percentage
FROM
    loan L
        INNER JOIN
    borrower B ON B.borrower_id = L.borrower_id
        INNER JOIN
(SELECT
    B.home_ownership AS own, COUNT(L.loan_id) AS count
FROM
    loan L
        INNER JOIN
    borrower B ON B.borrower_id = L.borrower_id
WHERE L.loan_status = "Fully Paid"
GROUP BY B.home_ownership) m
    ON m.own = B.home_ownership
GROUP BY B.home_ownership;
```

5C's credit analysis approach:

loan_status	loan_purpose	percentage
In Grace Period	Vacation	2.2222%
In Grace Period	Other	0.7160%
In Grace Period	Medical expenses	0.5917%
In Grace Period	Major purchase	0.4914%
In Grace Period	Home improvement	0.4281%
In Grace Period	Debt consolidation	0.3924%
Late (16-30 days)	Medical expenses	0.5917%
Late (16-30 days)	Debt consolidation	0.1308%
Late (16-30 days)	Home improvement	0.1223%
Late (16-30 days)	Other	0.1193%
Late (31-120 da...	Business	1.6529%
Late (31-120 da...	Major purchase	0.4914%
Late (31-120 da...	Debt consolidation	0.3779%
Late (31-120 da...	Other	0.3580%
Late (31-120 da...	Home improvement	0.1835%
Charged Off	Medical expenses	0.5917%
Charged Off	Debt consolidation	0.0872%
Charged Off	Home improvement	0.0612%

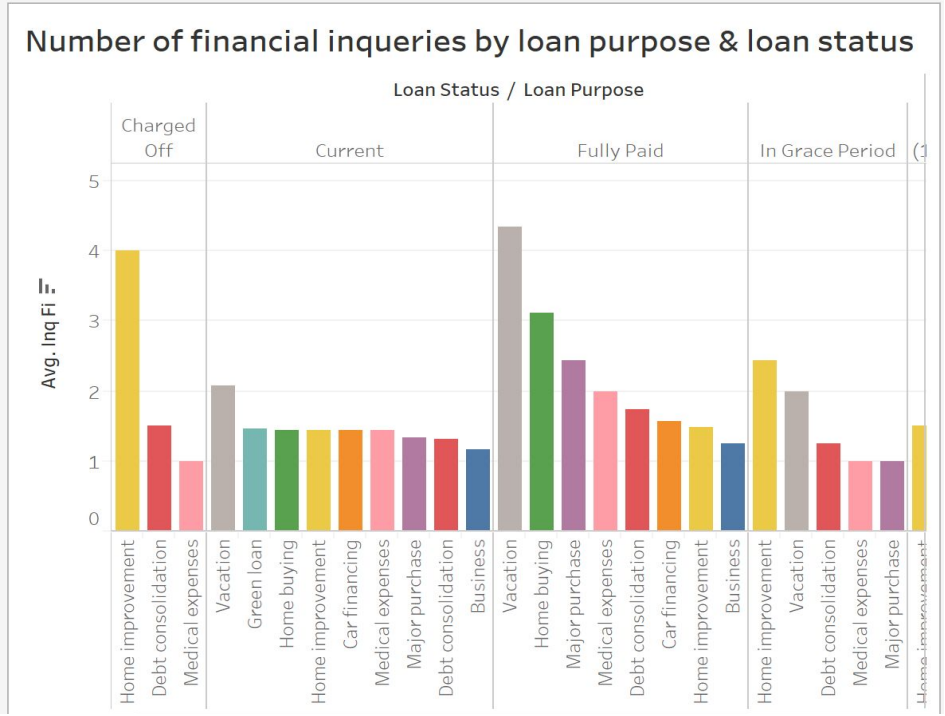
- Borrowers with which kinds of purposes more likely to default the loan?
- How was the loan grade given based on the information provided by borrowers.

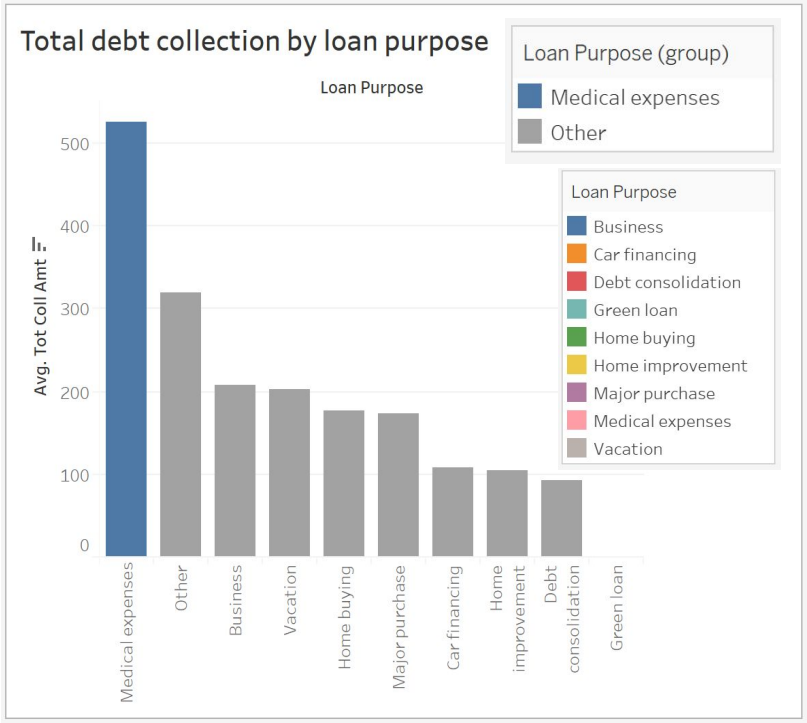
grade	annual_income	dti	delinquency	credit_line	revol_bal
A	100395.3244	19.930587	0.1301	26681.6080	18708.4211
B	97402.0686	22.878646	0.2229	26057.9445	19737.7978
C	93211.2430	22.676068	0.3291	25979.0026	20199.5197
D	90258.0065	22.948839	0.2839	24767.7419	18220.8581

```sql
# Q3: Borrowers with which kinds of purpose more likely to default the loan?

SELECT
    l.loan_status,
    t.loan_purpose,
    CONCAT(COUNT(l.loan_id)/count.count_all * 100, '%') AS percentage
FROM
    loan l
        INNER JOIN
    title_purpose t
            ON l.title_purpose_id = t.title_purpose_id
        INNER JOIN
    (SELECT
        t.loan_purpose AS purpose,
        COUNT(l.loan_id) AS count_all
    FROM
        loan l
            INNER JOIN
        title_purpose t
                ON l.title_purpose_id = t.title_purpose_id
    GROUP BY
        t.loan_purpose) count
            ON count.purpose = t.loan_purpose
    WHERE
        loan_status = 'charged off' OR
        loan_status like '%late%' OR
        loan_status = 'in grace period'

 GROUP BY
     t.loan_purpose,
     l.loan_status
 ORDER BY
     CASE    WHEN loan_status = 'in grace period' THEN '1'
             WHEN loan_status = 'late (16-30 days)' THEN '2'
             WHEN loan_status = 'late (31-120 days)' THEN '3'
             ELSE loan_status END ASC,
     percentage DESC;
```

```sql
# Q4: How was the loan grade given based on the information provided  by borrowers.
SELECT
    'A' AS grade,
    AVG(b.annul_inc) AS annual_income,
    AVG(b.dti) AS dti,
    AVG(b.delinq_2yrs) delinquency,
    AVG(b.total_acc)*1000 AS credit_line,
    AVG(b.revol_bal) AS revol_bal
FROM
    sub_grade s
        INNER JOIN
    loan l
            ON s.sub_grade_id = l.sub_grade_id
        INNER JOIN
    borrower b
            ON l.borrower_id = b.borrower_id
WHERE
    sub_grade like '%A%'

UNION ALL

SELECT
    'B' AS grade,
    AVG(b.annul_inc) AS annual_income,
    AVG(b.dti) AS dti,
    AVG(b.delinq_2yrs) delinquency,
    AVG(b.total_acc)*1000 AS credit_line,
    AVG(b.revol_bal) AS revol_bal
FROM
    sub_grade s
        INNER JOIN
    loan l
            ON s.sub_grade_id = l.sub_grade_id
        INNER JOIN
    borrower b
            ON l.borrower_id = b.borrower_id
WHERE
    sub_grade like '%B%'

UNION ALL

SELECT
    'C' AS grade,
    AVG(b.annul_inc) AS annual_income,
    AVG(b.dti) AS dti,
    AVG(b.delinq_2yrs) delinquency,
    AVG(b.total_acc)*1000 AS credit_line,
    AVG(b.revol_bal) AS revol_bal
FROM
    sub_grade s
        INNER JOIN
    loan l
            ON s.sub_grade_id = l.sub_grade_id
        INNER JOIN
    borrower b
            ON l.borrower_id = b.borrower_id
WHERE
    sub_grade like '%C%'

UNION ALL

SELECT
    'D' AS grade,
    AVG(b.annul_inc) AS annual_income,
    AVG(b.dti) AS dti,
    AVG(b.delinq_2yrs) delinquency,
    AVG(b.total_acc)*1000 AS credit_line,
    AVG(b.revol_bal) AS revol_bal
FROM
    sub_grade s
        INNER JOIN
    loan l
            ON s.sub_grade_id = l.sub_grade_id
        INNER JOIN
    borrower b
            ON l.borrower_id = b.borrower_id


WHERE
    sub_grade like '%D%';
```

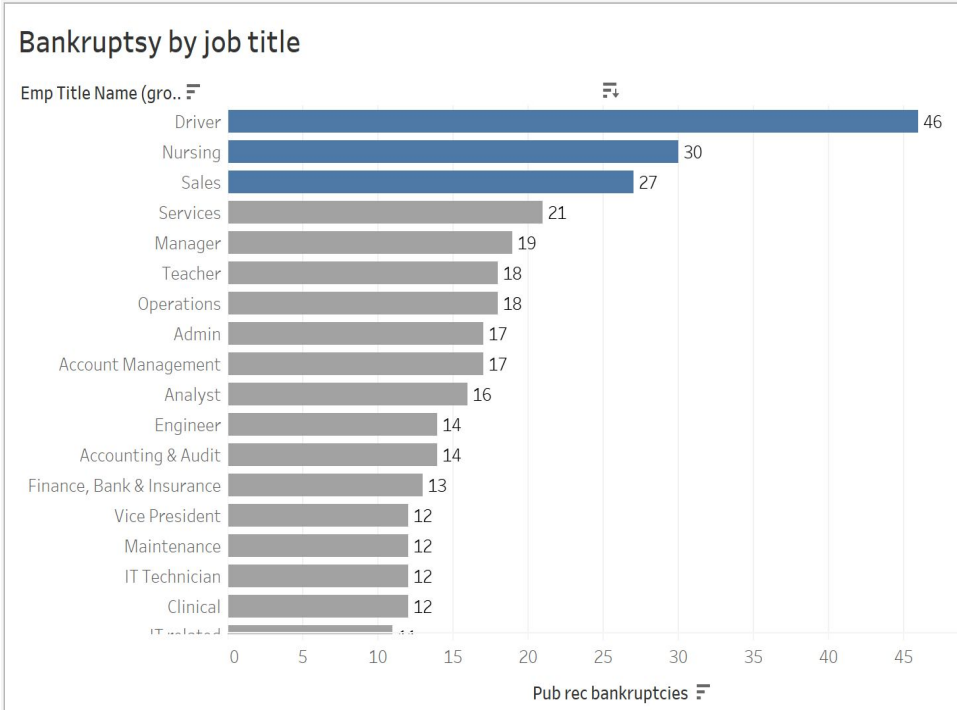
5C’s credit analysis approach:

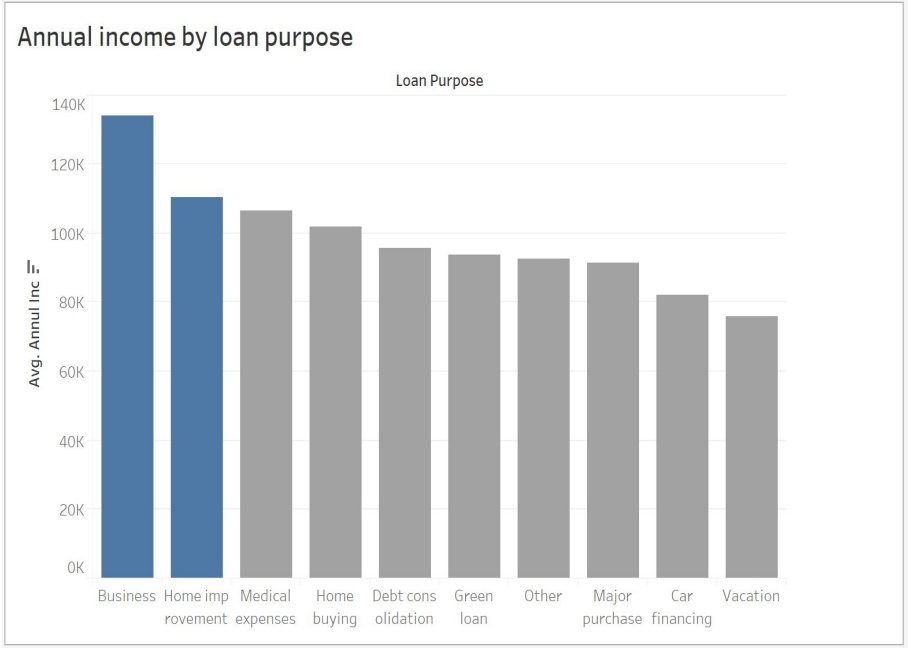




- **“Home Improvement”,”Vacation” and “Business”** incur most credit inquiries
- In charged off, grace period, and late related columns**, “Medical Expenses” and “Debt consolidation”** also pop up
- Moreover, “**Medical Expenses**” are more likely to become debt in collections

5C’s credit analysis approach:





- High income population are more likely to invest in **personal career development and life quality improvement**
- **Lower income and lower asset debtors** are more likely to go broke

Thank you!
Q&A