

# LENDING CLUB

---

A CASE STUDY ON RISK ANALYSIS

# CONSUMER FINANCE COMPANIES

---

- Risks involved:
  - Granting loans to everyone and suffering non-repayment of loans by some borrowers.
  - Not granting loans to people highly likely to repay their loans, hence losing out on business.

# DATA ANALYSIS

---

- Using data analysis tools in Python, the data of the applicants can be obtained and analysed.
- The main data-set here is the loan status.
  - The loan status of a person classifies them into: 'Charged off', 'Fully Paid' and 'Current'
- The data from a period of 2007 – 2011 is analysed.

# METHOD OF ANALYSIS

---

- Upon analysing the loan.csv file containing the data of the applicants, it is observed that:
  - Many columns have no data for any of the applicants.
  - Some columns have the same data for all of the applicants.
  - These columns are omitted
    - This is because the data in these columns don't distinguish between the applicants.
    - This makes it difficult for any meaningful analysis to be made.

# DATA CLEANING

---

- The dropping of columns with same data or no data throughout the entire list of applicants is an act of Data Cleaning.
- Furthermore, those columns with data of less than 30% of the applicants are also dropped.
- From amongst the remaining columns, some data need to be further cleaned
- Columns supposed to have numerical data, but has non-numerical characters in them are manipulated to drop their non-numerical characters.
- Year and Month from the column containing date are extracted and stored in new columns
- For some of the important datasets like Annual income, Interest rates, Loan amount and dti, ranges of values are made



# UNIVARIATE ANALYSIS

---

- The columns are put through univariate analysis, using the describe() function.
- The following are a few of the results:

```
#Univariate analysis - statistics of loan amount
```

```
loan['loan_amnt'].describe()
```

```
count    39717.00
mean     11219.44
std       7456.67
min        500.00
25%       5500.00
50%      10000.00
75%      15000.00
max      35000.00
Name: loan_amnt, dtype: float64
```

```
#Univariate analysis - statistics of annual income
```

```
loan["annual_inc"].describe()
```

```
count    39717.00
mean     68968.93
std      63793.77
min      4000.00
25%     40404.00
50%     59000.00
75%     82300.00
max     600000.00
Name: annual_inc, dtype: float64
```

# UNIVARIATE ANALYSIS

---

```
#Univariate analysis - statistics of interest rate
```

```
loan['int_rate'].describe()
```

```
count    39717.00  
mean       12.02  
std        3.72  
min        5.42  
25%       9.25  
50%      11.86  
75%      14.59  
max       24.59  
Name: int_rate, dtype: float64
```

```
#Univariate analysis - statistics of dti
```

```
loan['dti'].describe()
```

```
count    39717.00  
mean       13.32  
std        6.68  
min        0.00  
25%       8.17  
50%      13.40  
75%      18.60  
max       29.99  
Name: dti, dtype: float64
```

# UNIVARIATE ANALYSIS

```
#univariate analysis - loan amount statistics
```

```
loan['loan_amnt_categories'].describe()
```

```
count          39717
unique           7
top      5000-10000
freq           12960
Name: loan_amnt_categories, dtype: object
```

```
#univariate analysis - annual income statistics
```

```
loan['annual_inc_categories'].describe()
```

```
count          39717
unique           5
top      40000-60000
freq           11608
Name: annual_inc_categories, dtype: object
```

```
#univariate analysis - interest rate statistics
```

```
loan['int_rate_categories'].describe()
```

```
count          39717
unique           6
top           10-13
freq           12029
Name: int_rate_categories, dtype: object
```



# UNIVARIATE ANALYSIS

---

```
#Univariate analysis - statistics of state of residence
```

```
loan['addr_state'].describe()
```

```
# CA - California has the highest number of applicants
```

```
count      39717  
unique       50  
top         CA  
freq       7099  
Name: addr_state, dtype: object
```

```
#univariate analysis - dti statistics
```

```
loan['dti_categories'].describe()
```

```
count      39534  
unique        6  
top      10-15  
freq       9899  
Name: dti_categories, dtype: object
```

# UNIVARIATE ANALYSIS

---

```
#univariate analysis - statistics of 'purpose'
```

```
loan['purpose'].describe()
```

```
count          39717
unique           14
top      debt_consolidation
freq           18641
Name: purpose, dtype: object
```

```
#univariate analysis - statistics of 'home_ownership'
```

```
loan['home_ownership'].describe()
```

```
count      39717
unique         5
top         RENT
freq       18899
Name: home_ownership, dtype: object
```

# BIVARIATE ANALYSIS

---

The following are the plots of various criteria against the loan status.

Loan status vs Annual income of the applicant

annual_inc_categories	Charged Off	Current	Fully Paid	Total	Chargedoff_Proportion	Fully_paid_Proportion
0-20000	237	9	943	1189	0.20	0.79
20000-40000	1514	170	7004	8688	0.17	0.81
40000-60000	1729	345	9534	11608	0.15	0.82
60000-80000	1024	240	6597	7861	0.13	0.84
80000-6000000	1123	376	8872	10371	0.11	0.86

# BIVARIATE ANALYSIS

---

- Observations:
- The maximum proportion of the number of applicants charged off is in the category of annual income 0 to 20000, minimum is 80000 - 6000000
- The minimum proportion of the number of applicants fully paid is in the category of 0 to 20000, maximum is 80000 - 6000000.
- Therefore annual income is a good indicator of loan repayability in the future.
- The more the annual income, the more they are likely to pay it back.



# BIVARIATE ANALYSIS

## Loan status vs Purpose

loan_status	purpose	Charged Off	Current	Fully Paid	Total	Chargedoff_Proportion	Fully_Paid_Proportion
11	small_business	475.00	74.00	1279.00	1828.00	0.26	0.70
10	renewable_energy	19.00	1.00	83.00	103.00	0.18	0.81
3	educational	56.00	0.00	269.00	325.00	0.17	0.83
9	other	633.00	128.00	3232.00	3993.00	0.16	0.81
8	moving	92.00	7.00	484.00	583.00	0.16	0.83
5	house	59.00	14.00	308.00	381.00	0.15	0.81
7	medical	106.00	12.00	575.00	693.00	0.15	0.83
2	debt_consolidation	2767.00	586.00	15288.00	18641.00	0.15	0.82
12	vacation	53.00	6.00	322.00	381.00	0.14	0.85
4	home_improvement	347.00	101.00	2528.00	2976.00	0.12	0.85
1	credit_card	542.00	103.00	4485.00	5130.00	0.11	0.87
0	car	160.00	50.00	1339.00	1549.00	0.10	0.86
6	major_purchase	222.00	37.00	1928.00	2187.00	0.10	0.88
13	wedding	96.00	21.00	830.00	947.00	0.10	0.88



# BIVARIATE ANALYSIS

---

- Observations:
- The maximum proportion of the number of applicants charged off is in the category of 'Small Business', minimum is During a major purchase, Car loans and Wedding loans.
- The minimum proportion of the number of applicants fully paid is in the category of 'Small business', maximum is 'Major purchase' and 'Wedding'.
- Small businesses seem to have the highest risk, hence it is best to charge them higher rates of interest.

# BIVARIATE ANALYSIS

## Loan status vs Loan amount

loan_status	loan_amnt_categories	Charged Off	Current	Fully Paid	Total	Chargedoff_Proportion	Fully_paid_Proportion
6	30000-35000	180	93	555	828	0.22	0.67
5	25000-30000	143	85	557	785	0.18	0.71
4	20000-25000	542	163	2294	2999	0.18	0.76
3	15000-20000	751	242	3598	4591	0.16	0.78
0	0-5000	1314	96	8158	9568	0.14	0.85
2	10000-15000	1055	303	6628	7986	0.13	0.83
1	5000-10000	1642	158	11160	12960	0.13	0.86

# BIVARIATE ANALYSIS

---

- Observations:
- The maximum proportion of the number of applicants charged off is in the category of 30000 - 350000, minimum is 5000 - 10000
- The minimum proportion of the number of applicants fully paid is in the category of 30000 - 35000, maximum is 5000 - 10000
- According to a higher value of their fully paid proportion and a lower value of their charged off proportion, the loan amount ranges can be charged a lower or higher interest rate.
- It is interesting to note here that the most solvent group is in the neither too high nor too low range of loan amount

# BIVARIATE ANALYSIS

## Loan status vs Grade

loan_status	grade	Charged Off	Current	Fully Paid	Total	Chargedoff_Proportion	Fully_Paid_Proportion
6	G	101	17	198	316	0.32	0.63
5	F	319	73	657	1049	0.30	0.63
4	E	715	179	1948	2842	0.25	0.69
3	D	1118	222	3967	5307	0.21	0.75
2	C	1347	264	6487	8098	0.17	0.80
1	B	1425	345	10250	12020	0.12	0.85
0	A	602	40	9443	10085	0.06	0.94

# BIVARIATE ANALYSIS

---

- Observations:
- The maximum proportion of the number of applicants charged off is in the category of 'G', minimum is 'A'
- The minimum proportion of the number of applicants fully paid is in the category of 'G' and 'F', maximum is 'A'
- Here the grade is a very good indicator of the applicant's solvency.
- Higher Interest rates can be charged according to how bad their grades are.
- Also lending could be denied to applicants with a grade of F or lower.



# BIVARIATE ANALYSIS

---

Loan status vs dti

loan_status	dti_categories	Charged Off	Current	Fully Paid	Total	Chargedoff_Proportion	Fully_paid_Proportion
4	20-25	1118	237	5460	6815	0.16	0.80
3	15-20	1389	284	7422	9095	0.15	0.82
2	10-15	1402	269	8228	9899	0.14	0.83
5	25-30	87	53	536	676	0.13	0.79
1	5-10	1005	199	6868	8072	0.12	0.85
0	0-5	597	93	4287	4977	0.12	0.86

# BIVARIATE ANALYSIS

---

- Observations:
- The maximum proportion of the number of applicants charged off is in the category of 20 - 25, minimum is 0 - 10
- The minimum proportion of the number of applicants fully paid is in the category of 25 - 30, maximum is 0 - 5
- Here again the dti seems to be a good indicator of the applicant's solvency.
- Lower the dti, more likely the applicant is to repay the loan on time.
- Higher interest rates can be charged for applicants with a higher dti.

# BIVARIATE ANALYSIS

---

## Loan status vs Interest rate

loan_status	int_rate_categories	Charged Off	Current	Fully Paid	Total	Chargedoff_Proportion	Fully_Paid_Proportion
5	20-25	296	102	474	872	0.34	0.54
4	16-20	1266	359	3372	4997	0.25	0.67
3	13-16	1644	255	7527	9426	0.17	0.80
2	10-13	1591	347	10091	12029	0.13	0.84
1	6-10	775	75	10195	11045	0.07	0.92
0	0-6	55	2	1291	1348	0.04	0.96

# BIVARIATE ANALYSIS

---

- Observations:
- The maximum proportion of the number of applicants charged off is in the category of 20-25, minimum is 0-6
- The minimum proportion of the number of applicants fully paid is in the category of 20-25, maximum is 0-6
- Interest rates seem to determine a person's ability to repay the loan.
- However, higher interest rates are also a solution to solving the problem of lending to applicants with a charged off record,
- It could therefore be considered that slightly higher interest rates can be charged for applicant belonging to groups not very likely to default in general.
- As their likelihood of non-repayment increases however, the interest rates can be increased substantially.

# BIVARIATE ANALYSIS

---

Loan status vs number of public record bankruptcies

loan_status	pub_rec_bankruptcies	Charged Off	Current	Fully Paid	Total	Chargedoff_Proportion	Fully_Paid_Proportion
2	2	2.00	2.00	3.00	7.00	0.29	0.43
1	1	366.00	37.00	1271.00	1674.00	0.22	0.76
3	Not Known	118.00	0.00	579.00	697.00	0.17	0.83
0	0	5141.00	1101.00	31097.00	37339.00	0.14	0.83



# BIVARIATE ANALYSIS

---

- Observations:
- The maximum proportion of the number of applicants charged off is in the category of 2, minimum is 0
- The minimum proportion of the number of applicants fully paid is in the category of 2, maximum is 0
- Hence the number of public record bankruptcies can be thought of as a good indicator of the applicant's ability to repay the loan
- The more the number of public record bankruptcies, the less likely they are to not repay the loan

# CONCLUSION

---

- Higher annual income, lower dti, lower number of public bankruptcies, and a higher grade seem to indicate that the applicant is very likely to repay the loan.
- Lower annual income, higher dti, a higher number of public bankruptcies and a lower a grade seem to indicate that the applicant is more likely than not to default on their loan repayments.
- Interest rate is a useful tool, should the company decide to do business with the applicants of the latter category.
- The company should also keep in mind that a higher interest rate charged could also lead to a higher likelihood of the borrower defaulting in their repayments.