

Loan Database for  **LendingClub**

In today's presentation...

- **Problem Statement**
- **Proposal**
- **Normalization Plan**
- **ETL Process**
- **Analytical Procedures**
- **Database Interactions**



A peer to peer lending company, in which investors provide funds for potential **loan borrowers** and investors earn a profit depending on the risk they take. Lending Club provides the "bridge" between investors and borrowers.

Problem Description

Lending Club has a lot of archived, unstructured data that need to be organized for easy access and analysis:

- Raw dataset: we used is a reduced sample set that contains 5,000 records of **loans issued through the 2008-2018**, and **97 attributes** covering different information such as **current loan status, loan owner demographic information, customer financial credit history, payments, collections, secondary applicants**, etc.
- This dataset contains **redundant copies of** information in many places:

Example:

- *member_id* ---> *employee title, employee length, home ownership, income*
- *zipcode* (first three numbers) <--> *address_state c*
- *credit_grade/sub_grade* ----> *interest rate* issued to each member

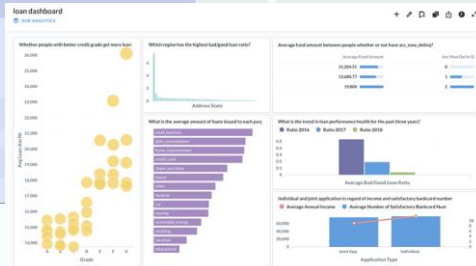
[illegible]

Field/View	Description
member_id	A Unique LC assigned id for the borrower member.
loan_id	A Unique LC assigned ID for the loan listing.
loan_bal	Maximum current balance owed on all revolving accounts
loan_status	Current status of the loan
loan_pymnt	The latest amount of the loan applied for by the borrower. If at some point in time, the credit department reduces the loan amount, then it will be reflected in this field.
loan_pymnt_d	Last month payment was received
loan_pymnt_amt	Last total payment amount received
code	The first 3 characters of the zip code provided by the borrower in the loan application.
verified_status_jan	Indicates if a two-borrowers' joint income was verified by LC, not verified, or if the income source was verified
verified_status_feb	Indicates if income was verified by LC, not verified, or if the income source was verified
title	The loan title provided by the borrower
term	The number of payments on the loan. Values are in months and can be either 36 or 60.
url	URL for the LC page with listing data.
rev_36_m	Total revolving high credit/credit limit
rev_gmp	Principal received to date
rev_pgm	Indicates if a payment plan has been put in place for the loan
gross	A category provided by the borrower for the loan request.
rev_bankruptcies	Number of public record bankruptcies
rev_credit_pull_d	The most recent month LC pulled credit for this loan
rate	The month which the loan was funded.
rate	Interest Rate on the loan
rev_ownership	The home ownership status provided by the borrower during registration or obtained from the credit report. Our values are RENT, OWN, MORTGAGE
rev_rate	Rate has received to date
rev_rate_int	Interest received to date
rev_pymnt_inv	Payments received to date for portion of total amount funded by investors
rev_pymnt_inv	Payments received to date for total amount funded
at_8_high_credit	Total installment high credit/credit limit
cu_8	Number of finance trades
rev_36_m	Total bankrupt high credit/credit limit
bal_d	Total current balance of all installment accounts

1. Loan Relational DBMS

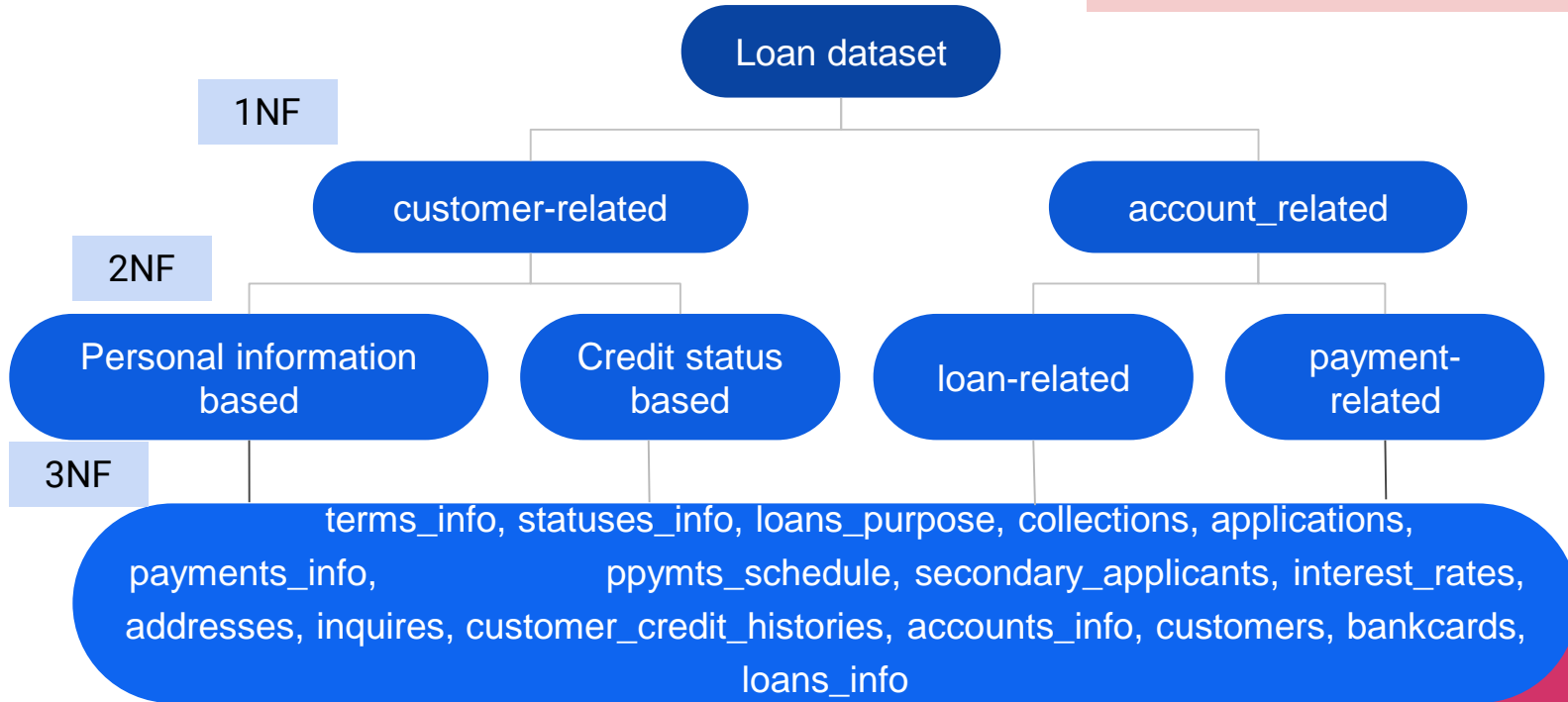
- Reduce redundancy with reorganized independent tables
- Achieve data consistency through PK/FK, constraints
- Enable triggers to enforce business rules that automatically perform an action after a change is made to database: *such to ensure each member inserted into the database fall into A,B,C,D,E,F,G categories of credit*
- Support simple structured query for daily operations
- facilitate business decision making

- Current Business Analysis
- Customer Analysis
- Risk Analysis



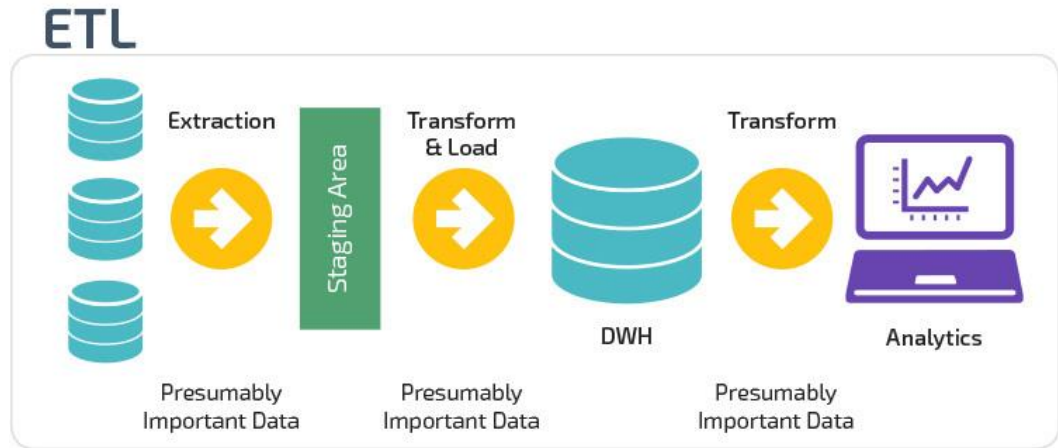
Normalization

1NF: include each table cell should contain a single value and each record needs to be unique
2NF: no non-prime attribute is dependent on the proper subset of any candidate key of table
3NF: to has no transitive functional dependencies

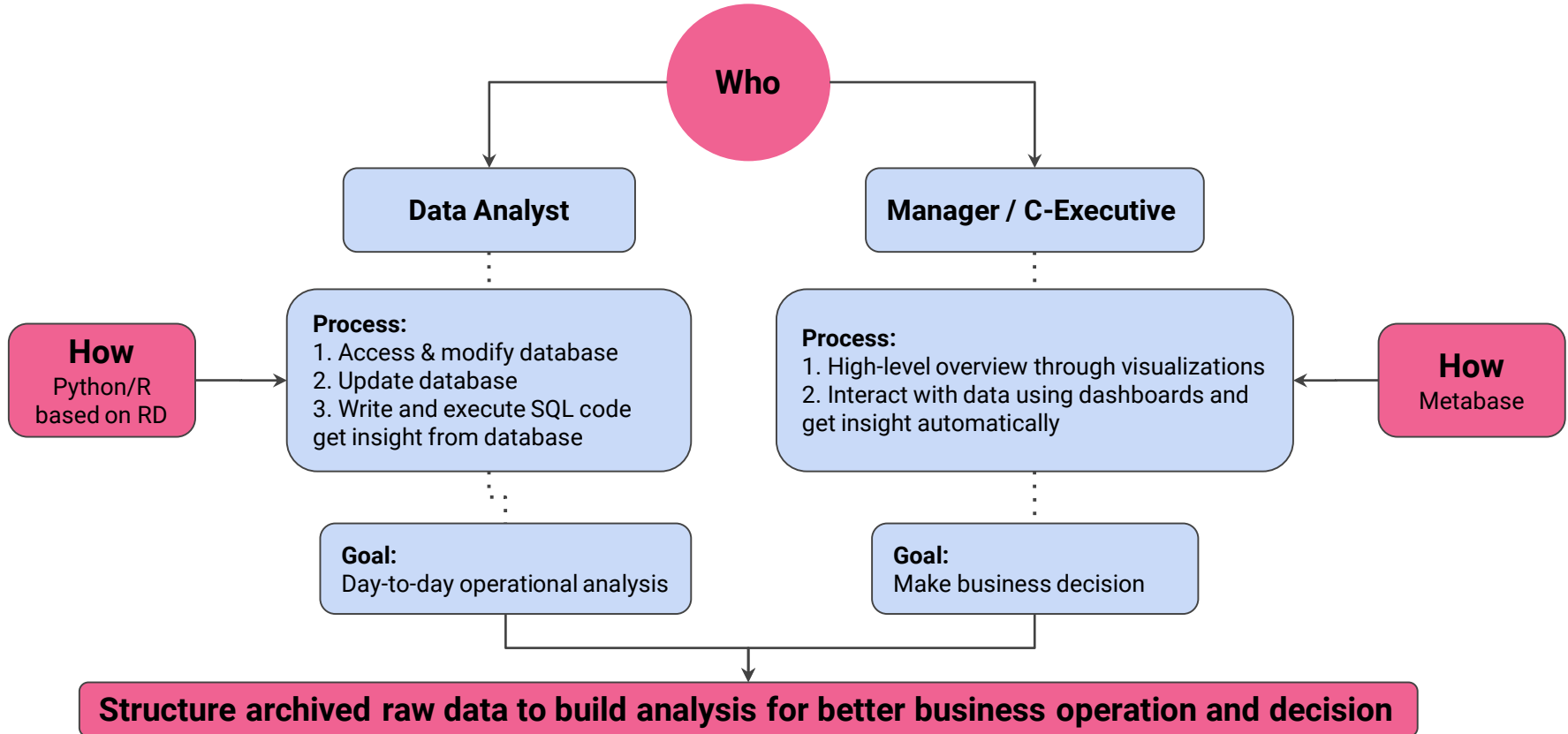


ETL Process

- Extract, Transform and Load.
- Moving the data into a data warehouse.
- In the extraction step, data is extracted from the source system into the staging area.
- In the transformation step, the data extracted from source is cleansed and transformed .
- In the loading data step, data is loaded into the warehouse .



Process of Interacting



Database Interaction with SQL : A DEMO

Is there a pattern for customers' credit history and funded loan amount?

Is there a difference between fund amount and customers with different count of public record of bankruptcy?

Query:

```
select avg(loan_amnt) as avg_loan_amnt, pub_rec_bankruptcies
from loans_info as L, customer_credit_histories as A, customers as C
where L.member_id = C.member_id AND C.credit_id = A.credit_id
group by A.pub_rec_bankruptcies
order by A.pub_rec_bankruptcies
```

	avg_fund_amnt numeric	pub_rec_bankruptcies integer
1	15493.863019891501	0
2	13032.398897058824	1
3	11788.392857142857	2
4	10666.6666666666666667	3
5	10800.0000000000000000	4

Is there a difference among fund amount and customers who have higher late fee and who have lower?

Query:

```
select avg(funded_amnt) as avg_fund_amnt, total_rec_late_fee
from loans_info
group by total_rec_late_fee
order by total_rec_late_fee
```

	avg_fund_amnt numeric	total_rec_late_fee double precision
1	15194.402985074627	0
2	28000.0000000000000000	6.6e-09
3	7000.000000000000000000	2.3
4	9000.000000000000000000	7
5	1200.000000000000000000	14.9132948
6	2200.000000000000000000	14.93382961
7	1500.000000000000000000	14.97454607
8	10275.000000000000000000	14.99999994
9	7104.444444444444444444	15
10	2400.000000000000000000	15.00000001
11	4500.000000000000000000	15.00000004

Fig 6.2 Average Fund Amount with Late Fee

Is there a difference among fund amount and whether or not a customer has acc_now_delinq?

Query3:

```
select avg(funded_amnt) as avg_fund_amnt, acc_now_delinq
from loans_info as L, accounts_info as A, customers as C
where L.member_id = C.member_id AND C.account_info_id = A.account_info_id
group by acc_now_delinq
order by acc_now_delinq
```

	avg_fund_amnt numeric	acc_now_delinq integer
1	15204.512635379061	0
2	13680.769230769231	1
3	19800.0000000000000000	2

Fig 6.3 Average Fund Amount in Delinq Account

Summary: It is not likely that people with less acc_now_delinq to receive more average fund amount.

It is not likely that people with less late fee to receive more average amount.

People with more public record bankruptcies tend to receive less average fund amount

Database Interaction with Metabase: A DEMO

Accounts Info

Avg Loan By Grade

Customer Credit Histories

Inquires

Loans Purpose

Secondary Applicants

Addresses

Bank Cards

Customers

Interest Rates

Payments Info

Statuses Info

Applications

Collections

Customers Audits

Loans Info

Py Mn Ts Schedule

Terms Info

START HERE

Current Business Analysis

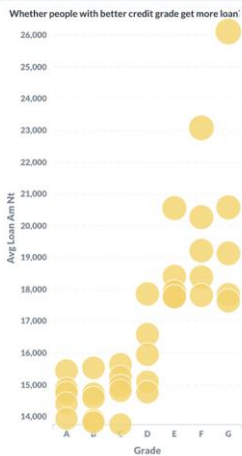
Customer Analysis

Risk Analysis

loan dashboard


OUR ANALYTICS

Whether people with better credit grade get more loan?




Grade	Avg Loan Amt
A	14,500
B	15,000
C	15,500
D	16,000
E	17,000
F	18,000
G	19,000

Which region has the highest bad/good loan ratio?



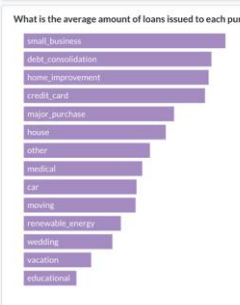
Address State	Bad/Good Ratio
small_business	5.5
debt_compensation	4.5
home_improvement	4.0
credit_card	3.5
major_purchase	3.0
house	2.5
other	2.0
medical	1.5
car	1.0
moving	0.5
renewable_energy	0.2
wedding	0.1
vacation	0.05
educational	0.02

Average fund amount between people whether or not have acc_now_delinq?



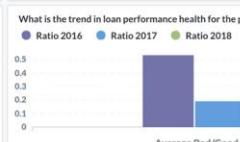
acc_now_delinq	Average Fund Amount
0	15,204.51
1	13,680.77
2	19,800

What is the average amount of loans issued to each purp



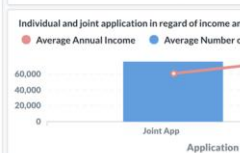
Purpose	Average Bad/Good Loan Ratio
small_business	0.55
debt_compensation	0.45
home_improvement	0.40
credit_card	0.35
major_purchase	0.30
house	0.25
other	0.20
medical	0.15
car	0.10
moving	0.05
renewable_energy	0.02
wedding	0.01
vacation	0.005
educational	0.002

What is the trend in loan performance health for the past three years?



Ratio	2016	2017	2018
Ratio 2016	0.45	0.25	0.15
Ratio 2017	0.25	0.15	0.10
Ratio 2018	0.15	0.10	0.05

Individual and joint application in regard of income and satisfactory bankcard number



Application Type	Average Annual Income	Average Number of Satisfactory Bankcard Num
Joint App	40,000	40,000
Individual	60,000	60,000

Our analytics

What is the loan approval rate for joint applications in 2018?

What is the average amount of loans issued to each purpose?

Whether people with better credit grade get more loan?

When should we suggest customers to apply for individual application or for joint application?

What is the difference between individual and joint application in regard of income and satisfactory bankcard number?

Which region has the highest bad/good loan ratio?

Where does the best market locate?

What is the trend in loan performance health for the past three years?

What is the average amount of loans issued to each purpose?

Bad/Good Ratio



Address State	Bad/Good Ratio
small_business	5.5
debt_compensation	4.5
home_improvement	4.0
credit_card	3.5
major_purchase	3.0
house	2.5
other	2.0
medical	1.5
car	1.0
moving	0.5
renewable_energy	0.2
wedding	0.1
vacation	0.05
educational	0.02

Flow Chart



Purpose	Average Bad/Good Loan Ratio
small_business	0.55
debt_compensation	0.45
home_improvement	0.40
credit_card	0.35
major_purchase	0.30
house	0.25
other	0.20
medical	0.15
car	0.10
moving	0.05
renewable_energy	0.02
wedding	0.01
vacation	0.005
educational	0.002

Conclusion: How will Loan RDMS benefit Lending Club?



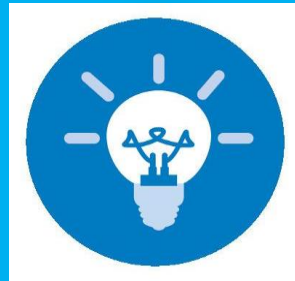
Analysts

The Loan RDMS allows analysts to write and execute SQL code to run day-to-day operational analysis and generate report for management.



C- executives

Metabase, connected with Loan RDMS, will provide visualized insights and interactive dashboards that automatically update when new data is inserted to database, making it straightforward for non-tech executives.



Overall

Loan RDMS helps Lending Club achieve a more smooth operation process and facilitate business decision-making.