Bank Statement Analysis

Oct-2018

Agenda

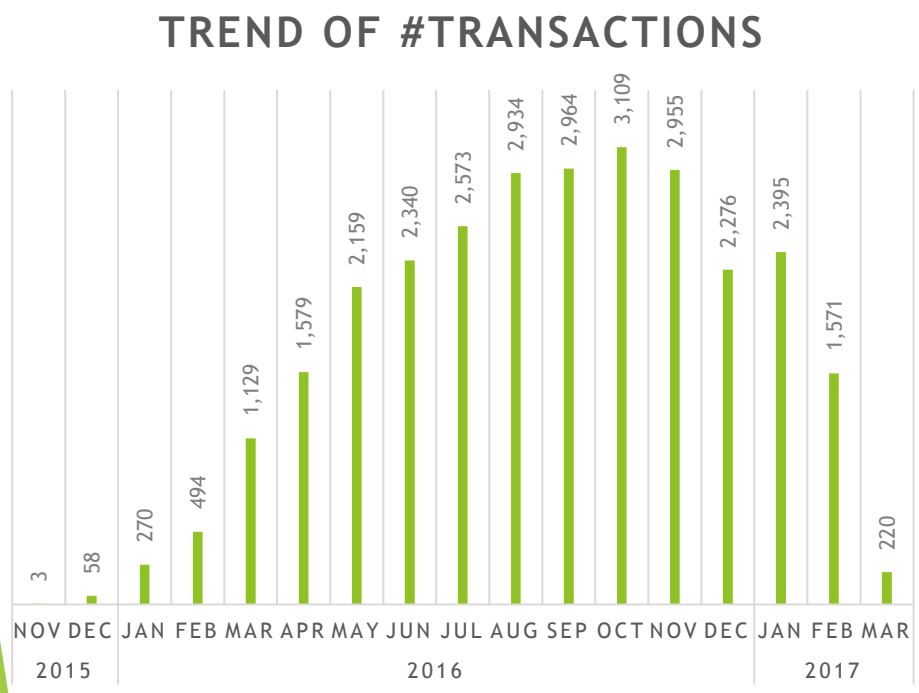
- Data Overview
- Trend of Transactions over time
- Merchant Analysis
  - Transactions and Amount
  - Average Transaction Amount
- Industry Level Analysis
- Solution to Exercises

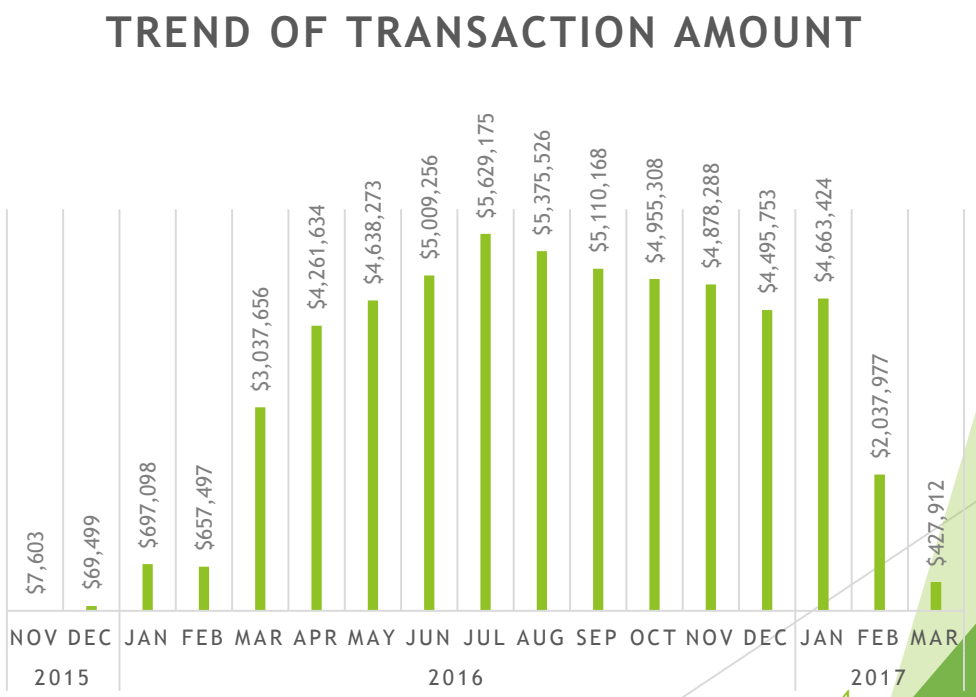
Data overview

- This is a sample of individual bank statement transactions from 21 small business merchants from various industries
- The time frame is 17 months, spanning over Q4-2015 to Q1-2017
- 29k transactions were made during this period amounting to ~$56M
  - 64% (18k) of these were debit transactions and rest were Credit transactions
  - The amount associated with these transactions was almost same (50% each)
- These 21 merchants represent 10 different Industries
- Most of these merchants deal with single bank with an average of 2 accounts
- Average credit transaction amount is $2.6k and for debit it's $1.5k
- Merchant 321146 has two account number for same bank account id in same bank.
  - Second account opened in Dec-2016 and heavy transactions since then

Trend of Transactions over time

- On an average there are 2k transactions with average transaction amount of $4M
- Data for 2015 and Mar-2017 seems incomplete

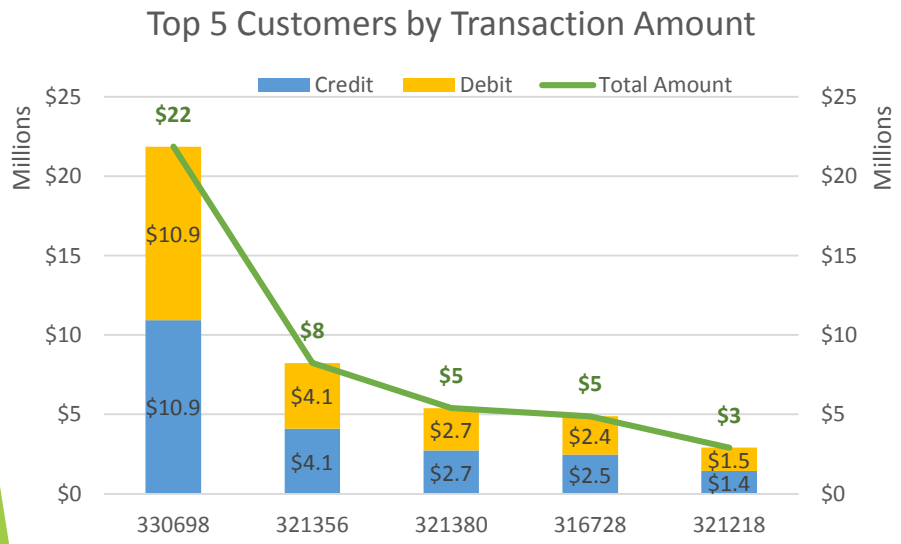


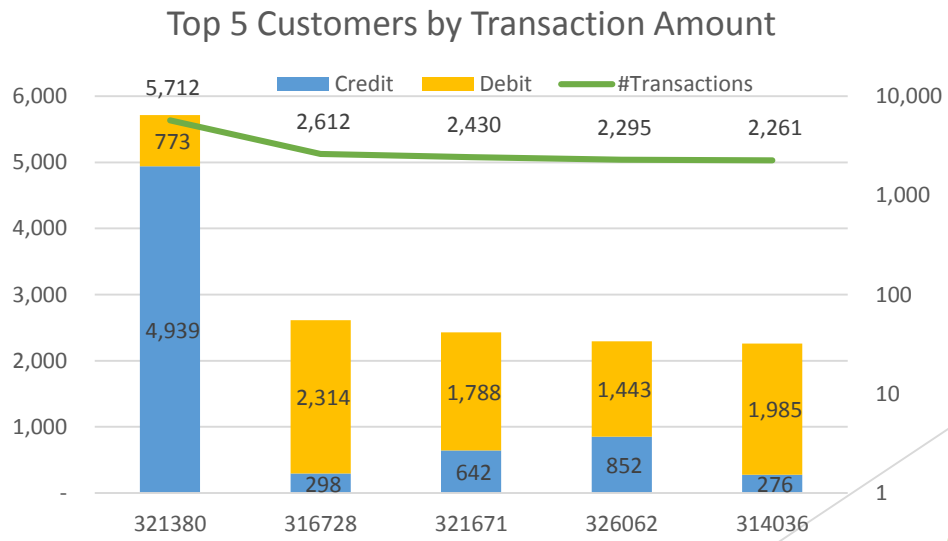


Merchant Level Analysis

Transactions and Amount

- Merchant 330698 has highest transaction amount ($22M/$56M)
  - It belongs to Health Care and Social Assistance industry
- Merchant 321380 has highest number of transactions
  - It has high proportion of Credit transactions which is opposite to other merchants

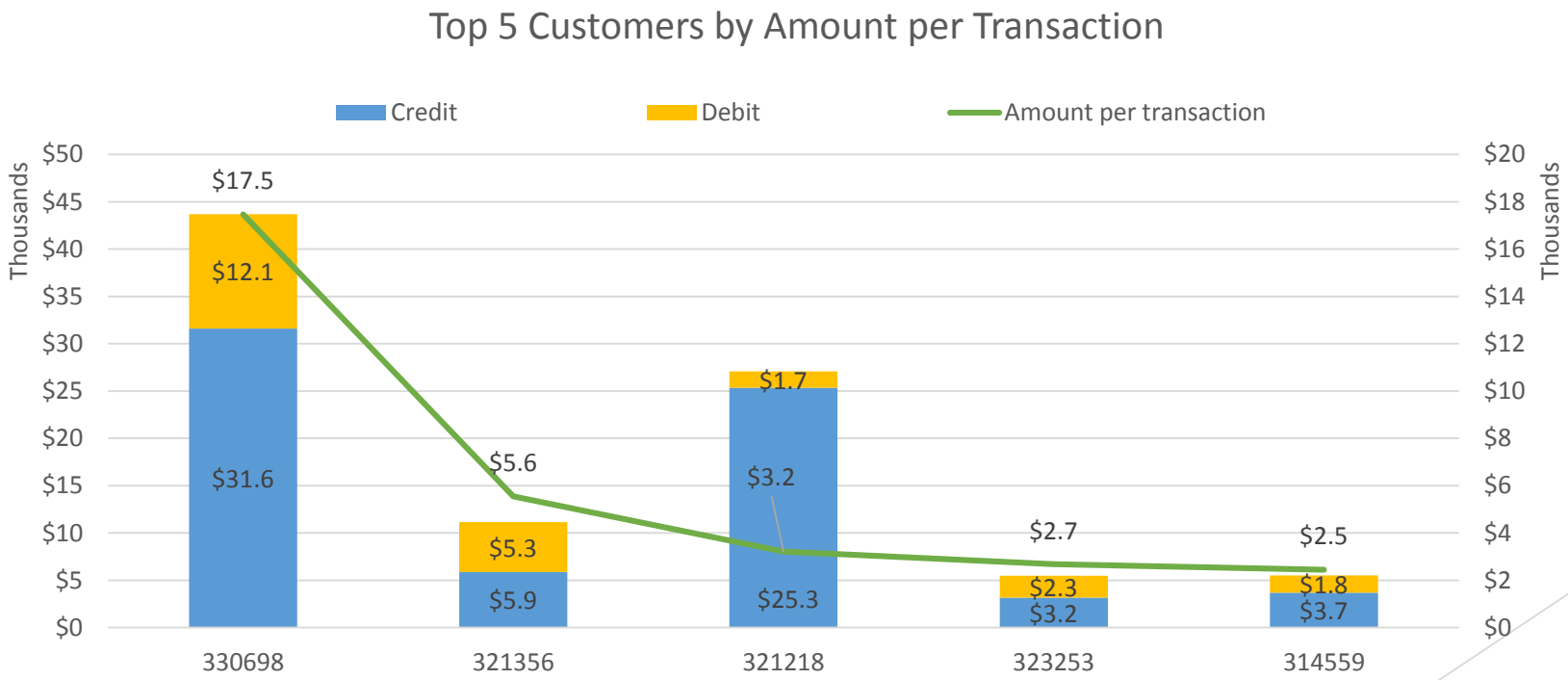




Merchant Level Analysis

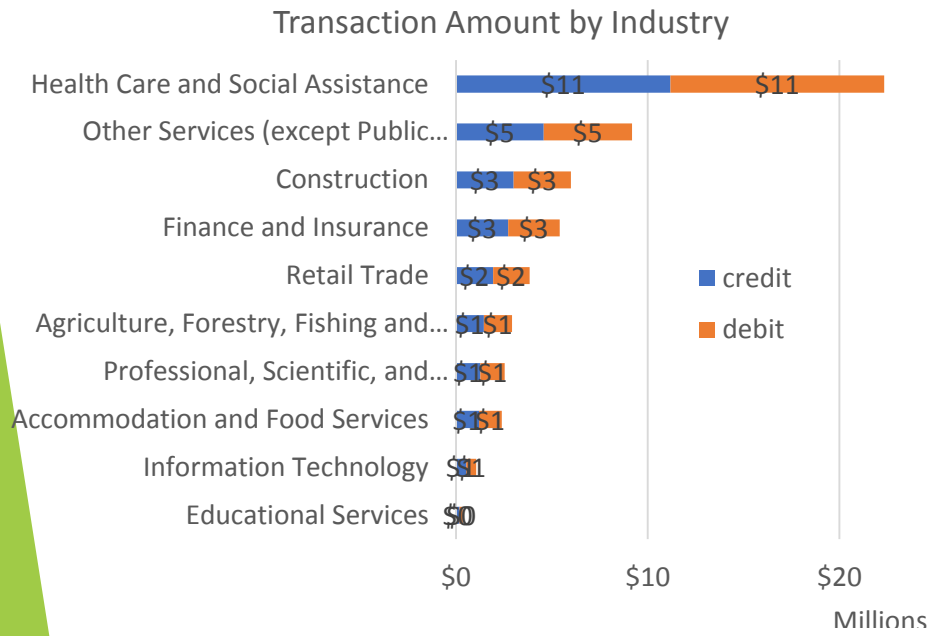
Average Transaction Amount

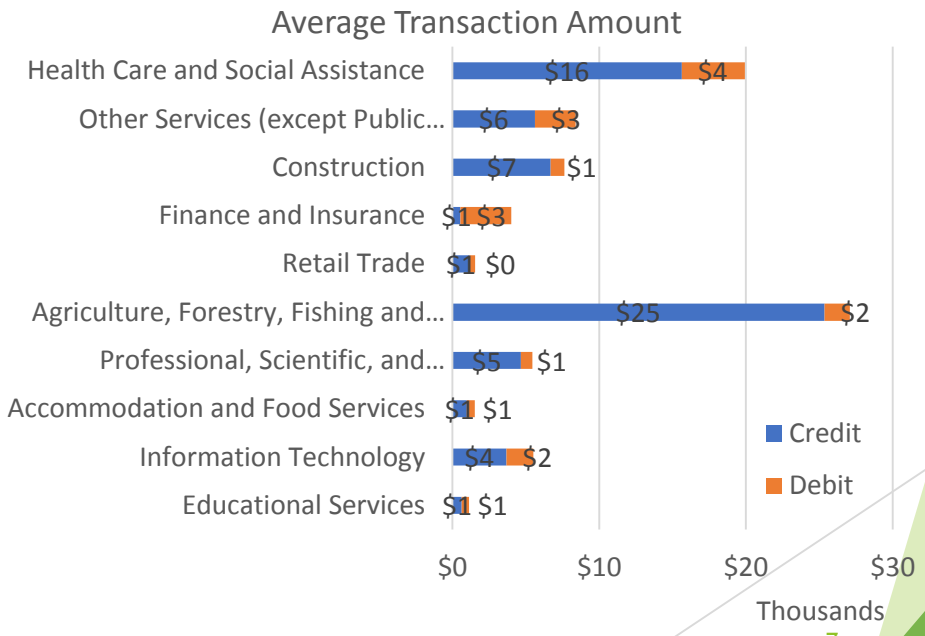
- Merchant 330698 have highest average transaction amount
  - It's average credit transaction amount is highest
- Merchant 321218 has quite high average credit transaction amount



Industry level Analysis

- Health care industry has highest transaction amount
  - This may be due to 1 merchant (330698)
- Agriculture industry has highest average transaction amount

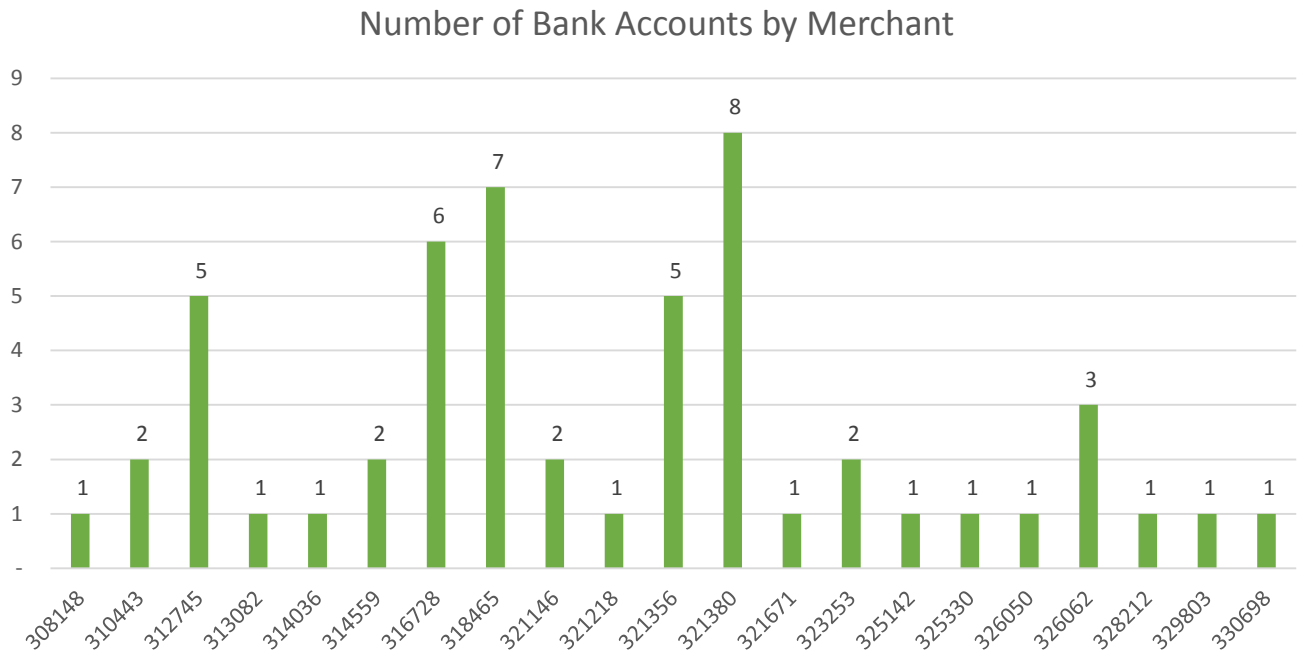




Solutions to Exercises

Exercise 1

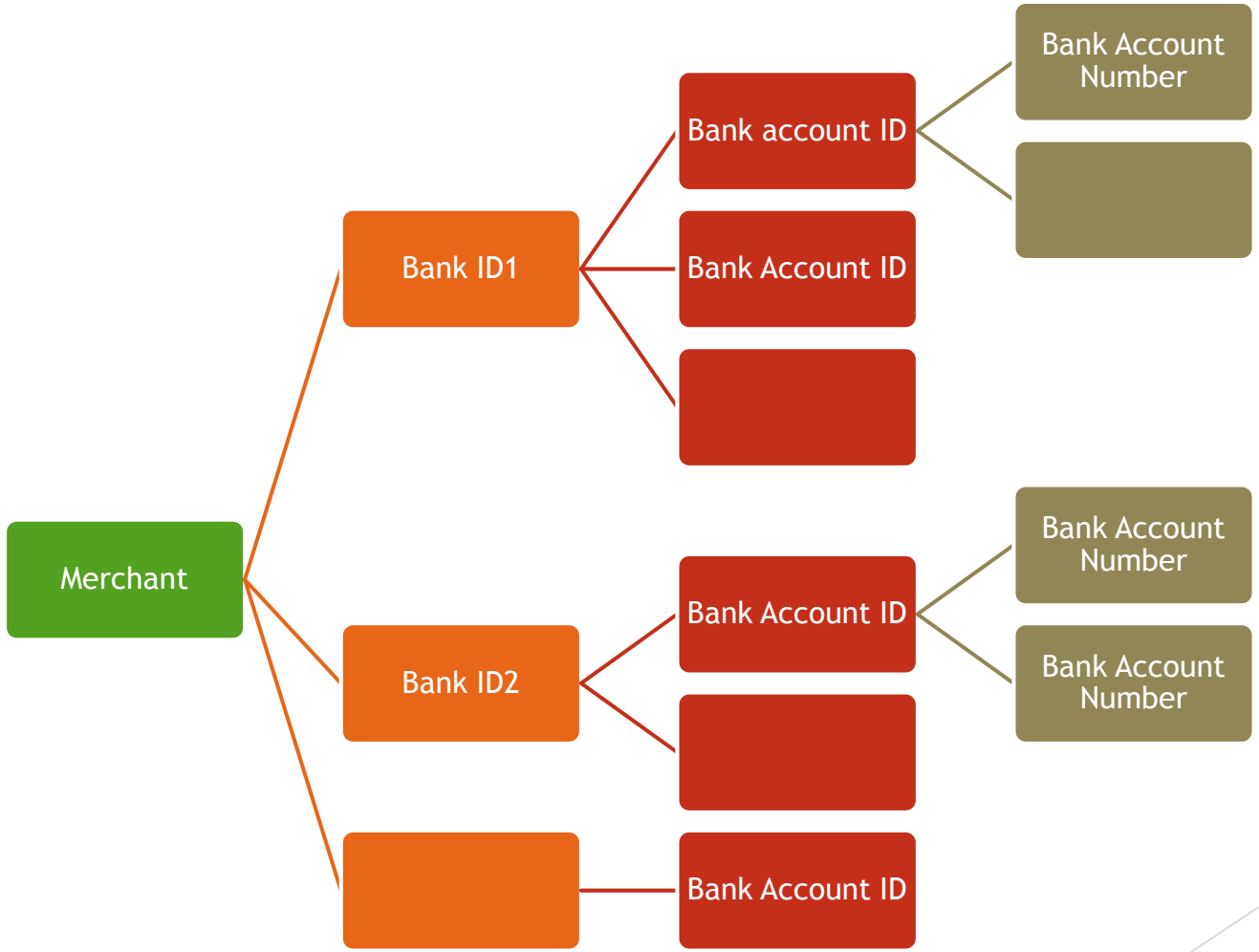
- (i) the number of bank accounts for each merchant
  - Each merchant has average of 3 bank accounts
- (ii) the number of months of each bank account for which data is available
  - Each merchant has data for an average of 9 months for each account



LeadID	bank_account_id	#Months with data
308148	12460	10
310443	12654	7
	12655	12
312745	12835	5
	12836	5
	12837	1
	12838	11
	12839	13
313082	13233	10
314036	13226	11
314559	13271	12
	13273	7
316728	13228	12
	13230	7
	13232	13
	13234	13
	13235	11
	13272	8
318465	13419	9
	13420	3
	13421	2
	13422	7
	13423	8
	13424	9
	13425	8
321146	13970	9
	13971	4
321218	14779	13
321356	15142	13
	15144	13
	15146	13
	15147	11
	15148	11
321380	14630	2
	14631	2
	14632	2
	14633	2
	14634	12
	14635	12
	14636	13
	14637	13
321671	14629	13
323253	13877	11
	13879	8
325142	14049	8
325330	15002	11
326050	14206	8
326062	14044	11
	14045	2
	14046	8
328212	14981	13
329803	14339	13
330698	14374	13

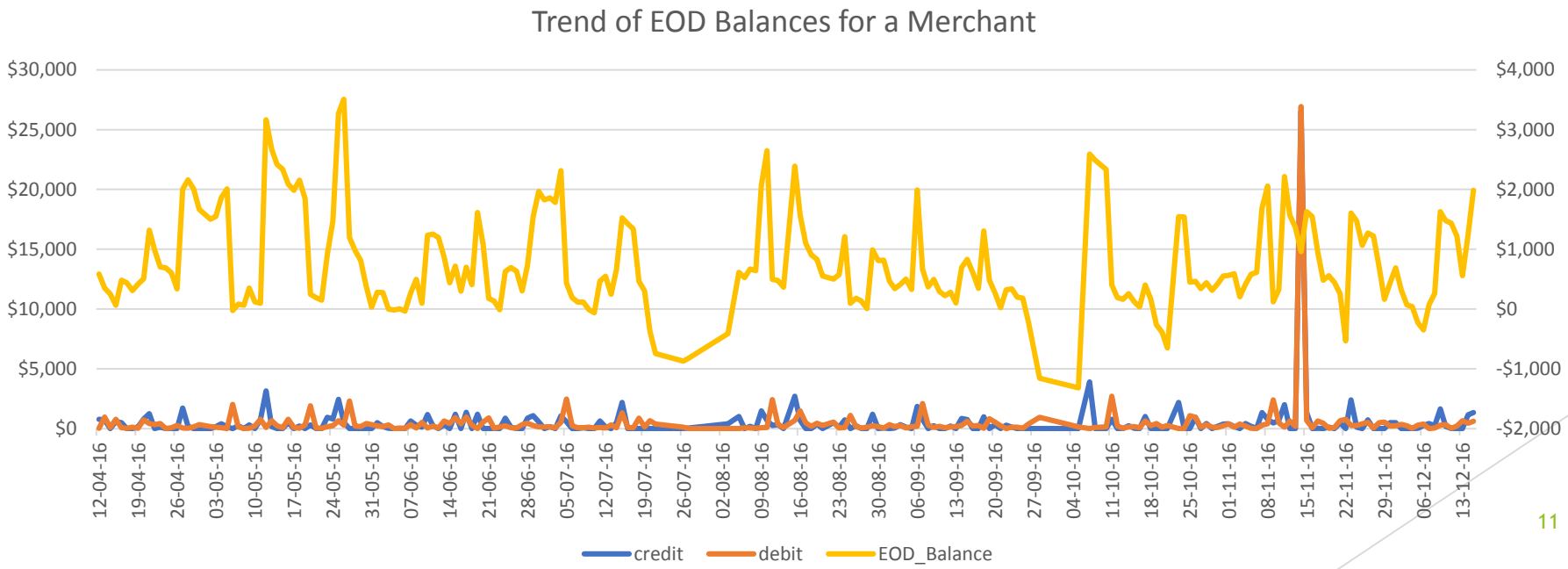
Exercise 1

- The relationship between LeadID, bankid, bank_account_id and account_number



Exercise 2

- The average EOD Balance is $746
- There is big withdrawl in 14-Nov-2016 ($27k) and deposit of $26k as well

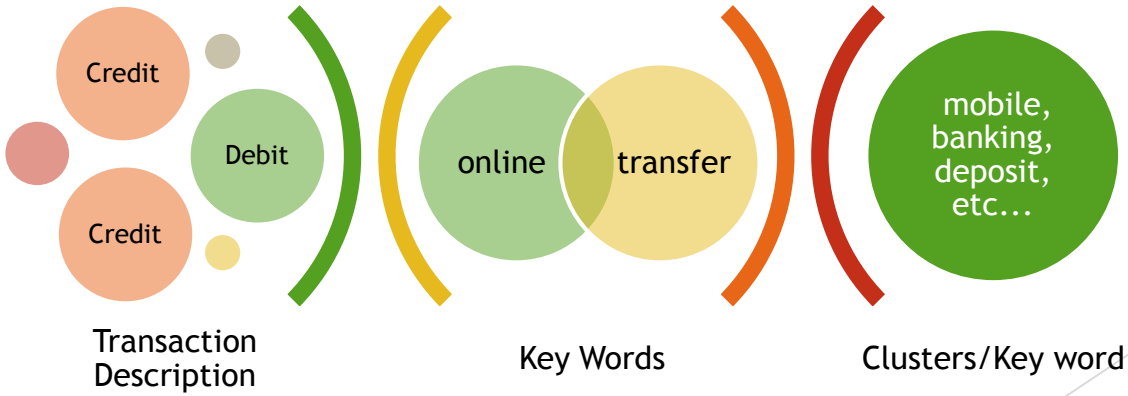


Some cash flow variables that could be relevant for predicting merchant default

- Number of times running balance is negative in last 12 months
- Number of months when the average daily withdrawl is more than average daily deposits
- Trends in EOD balance, is it increasing or decreasing recently
- Variation in amount of deposits/ withdrawls i.e., consistent transaction of average amounts or few high value transactions
- Source/ mode of transactions

Exercise 5. Develop a program by which clusters of similar transactions for a merchant can be identified from transaction descriptions

- The description of each transaction of an Merchant was analyzed and grouped into 4 clusters
  - The credit and debit transactions were separately analyzed
  - The description was characterized by key words appearing in it
  - The transactions were grouped into clusters using above key words
  - Top key words of each cluster can be analyzed to define the groups for business



Cluster Analysis of description for the Merchant 318465 and Credit transactions

- The grouping of similar text can be done by K means clustering and Topic modeling. As the description is short, K-means algorithm preferable
- The analysis shows that there are 4 clusters and the top words representing these clusters are as below

Cluster 1	Cluster 2	Cluster 3	Cluster 4
• CK • Online • Transfer • Banking • Deposit	• Dividend • Amp • Amazon • Alderwood • Aker • Airbnb • Ahlo • Advocare	• Mobile • Banking • Deposit • Card • Ending • POS • Waus • Bothell	• External • Square • Deposit • Payments • Advocare • Stripe • Transfer • Stylehaul