Project Flow Diagram

1. Data Gathering:

Data Downloaded from Kaggle (Lending club data)

Clean data using R and created dummy Customer dataset

Û

3. Data Post -Processing:

Two hive tables

1. Customer 2. Loan data

Total 6 views have been created on hive based on two tables.

Hive query result to HDFS

- 1.HBase table created.
- 2. Data load to HBase

2. Data Pre-Processing

1.Remove Header and "" from the file. 2.Create file with "m_" for MySQL and "h_" for HDFS.

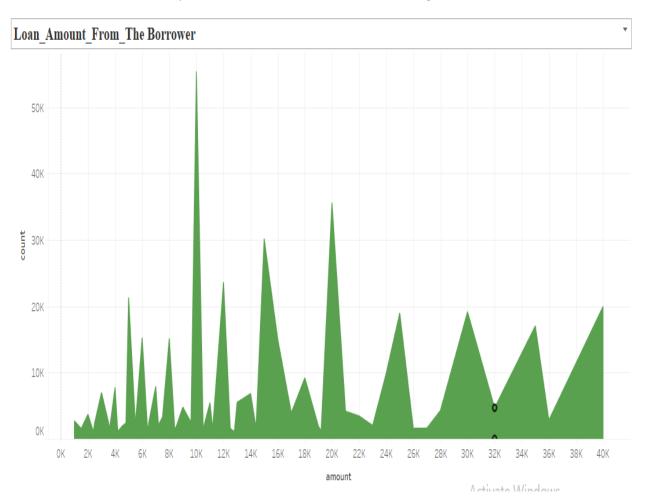
1.Run Maprduce.sh Script on "h_" 2. stored "m_" file in MySQL

Sqoop to create table and load data from MySQL-hive. 2.Create external table on the output of MapReduce job.

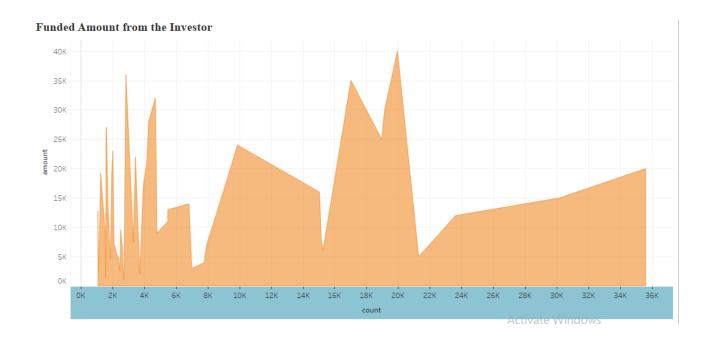
I. RESULT

All the Visualization created with the Reporting Tool Tableau.

1. Maximum loan amount requirement from the borrower is in the range of 20000-40000 USD.



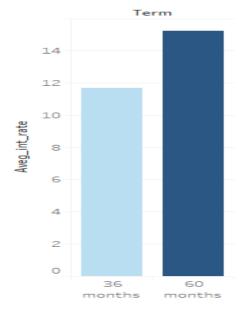
2. Maximum invest from the Investor is in the range of 20000 USD to 40000 USD. This means that potential borrower gets the loan whatever loan is applied by the borrower



3. average interest rate is important factor while purchasing the loan. in the dataset two column are present. Which are int rate and term. int rate define how much interest has been applied to the loan. and term is period for which loan is taken.

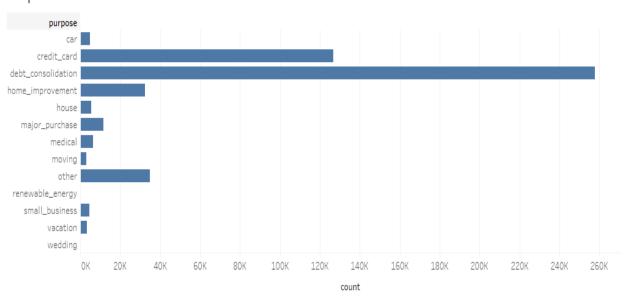
From the analysis for 36-month period average interest rate is minimum and for loan period which is 60 interest rates is 15.20363

Average Interest Rate in USA by Term



4. loan purpose column in the dataset contains the category of the loan purpose. From the analysis it is clear that most of the time reason of application is debt consolidation. After that maximum application has been created for the credit card. Lowest loan application is for the wedding and renewable energy

Purpose Of The Loan



5. Highest annual average income in the state Washington d.c. Idaho is the state where there is less amount of the annual average income. This analysis is important in business perspective.as this filed with other filed defines the average interest rate for the loan application from these state

Average Annual Income By State

DC 99,698	MA 83,774	CO 79,608	UT 76,168	LA 76,059		NV 75,768		AZ 75,466		NC 75,193	FL 74,603
MD 92,067	NY 82,536	ND 79,404	MI 74.080		TN 72,138		NM	24	KS	KY	WI 71.570
NJ 90,803	AK 82,277	NH 79,366	74,080 OK 73,989 SC 73,807		12,	138	71,92	24	71,784	1 71,78	2 71,578
VA 88,110	IL 81,731	DE 79,260			MO 71,573			R 0,227	ME 68,028	AR 67,596	
CA 86,794	WA 80,287	GA 79,247	WY 73,315		OH 71,304 MS			_			
CT 86,482	HI 79,734	MN 77,289	AL				6	SD 67,360 VT 67,048		MT 66,124	
TX	RI	PA	72,873 WV		71,028 IN						ID 65,968
83,885	79,693	76,592	72,330		70,857			E 6,351		Δ.	

6. Average interest rate is analyzed according to the state of the us. From the analysis it is clear that annual income affect the interest rate. by consider the low annual income state in USA ID, AR, VT the Annual interest rate for this state is high.

Average Interest Rate By State

HI 13.443	SD 12.980	SC 12.866	MI 12.795	OH 12.783		KY 12.777		M 2.765	WA 12.751		NV 12.737	NC 12.735
AL 13.236	GA 12.970	D state: 1; interest_rate	SC e: 12.866									
MS 13.210	CT 12.932	OK 12.859	PA 12.727		MO 12.6	OR 12.661		51	RI 1 12.654		NJ 12.640	TX 12.614
VT 13.088	ND 12.901	AK 12.827	IN 12.724	CA 12.610			UT	551	KS		ME 12.516	
AR 13.069	MD 12.893	NE 12.814	MT 12.700		IL 12.606			12.	12.551		510	12.316
NY 13.017	WV 12.889	TN 12.806	WY 12.684	AZ 12.577				MN 12.507		CO 12.442	MA 12.436	
ID 12.981	LA 12.886	FL 12.803	VA 12.680		DC 12.	DC 12.562		WI 12.	WI 12.481		NH 12.341	

A ~+i