

# Project Plan for Final Project

## Advanced Database Management

Submitted by Group 4

Uma Srikanth Reddy | Prasanna Eddala | Mohan Vamsi Karasala | Akhilasri Musikara

### Project Proposal:

We plan to design a Loan disbursement data set from U.S. Small Business Administration. The database once created could be used to answer the question: “should the loan be approved or denied?”.

### Project Plan:

The data from the excel worksheet will be loaded into Python for preprocessing. Then, tables will be created using Oracle SQL Developer and the data will be loaded into the database in accordance with the Entity Relationship Diagram. The ER diagram will be modified using the 1NF, 2NF, and 3NF standard forms to ensure data integrity and eliminate redundancy. During this process, data will be modeled using machine learning models in Python at the same time. The results of the machine learning models could be visualized using Tableau. As a result, we will be able to develop a portfolio project.

### Plan of Work:

Topic	Description	Deadline
Database Design	During this phase, the goal is to construct a conceptual and logical model of the database using ERD and Normalization, which accurately represents the requirements of the system and ensures the quality and integrity of the data.	March 26 <sup>th</sup> 2023
SQL Query Writing	Develop SQL queries to provide user suggestions based on the use cases specified and develop stored procedures in the database.	April 8 <sup>th</sup> 2023
Query and Transaction Performance	A query processor provides users with quick and efficient access to the information they need, and a transaction processor ensures the consistency and reliability of the data in the database.	April 22 <sup>nd</sup> 2023
Other Topics	Data Visualization and Providing Database Security	April 29 <sup>th</sup> 2023

3. The raw data set consists of 899164 records and 27 attributes. Each of the attributes before preprocessing of the data are listed below:

Variable name	Data type	Description of variable
LoanNr_ChkDgt	Text	Identifier – Primary key
Name	Text	Borrower name
City	Text	Borrower city
State	Text	Borrower state
Zip	Text	Borrower zip code
Bank	Text	Bank name
BankState	Text	Bank state
NAICS	Text	North American industry classification system code
ApprovalDate	Date/Time	Date SBA commitment issued
ApprovalFY	Text	Fiscal year of commitment
Term	Number	Loan term in months
NoEmp	Number	Number of business employees
NewExist	Text	1 = Existing business, 2 = New business
CreateJob	Number	Number of jobs created
RetainedJob	Number	Number of jobs retained
FranchiseCode	Text	Franchise code, (00000 or 00001) = No franchise
UrbanRural	Text	1 = Urban, 2 = rural, 0 = undefined
RevLineCr	Text	Revolving line of credit: Y = Yes, N = No
LowDoc	Text	LowDoc Loan Program: Y = Yes, N = No
ChgOffDate	Date/Time	The date when a loan is declared to be in default
DisbursementDate	Date/Time	Disbursement date
DisbursementGross	Currency	Amount disbursed
BalanceGross	Currency	Gross amount outstanding
MIS_Status	Text	Loan status charged off = CHGOFF, Paid in full = PIF
ChgOffPrinGr	Currency	Charged-off amount
GrAppv	Currency	Gross amount of loan approved by bank
SBA_Appv	Currency	SBA's guaranteed amount of approved loan

## Entity Relationship Diagram:

This is a logical ERD which is in non-normalized form.

