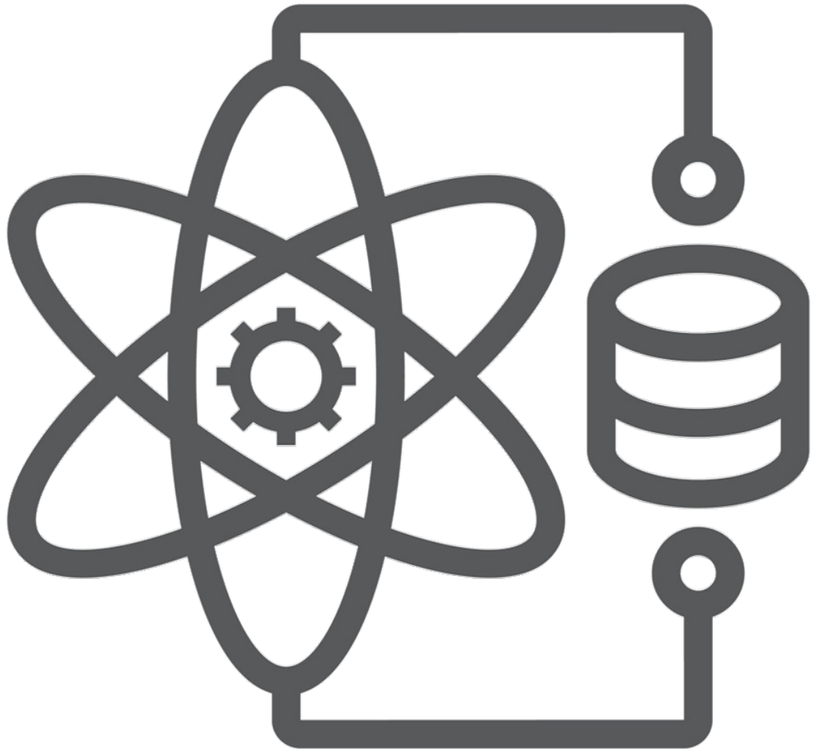
simplilearn | PURDUE UNIVERSITY

Data Science Capstone Project

Introduction to Capstone Projects

What Is DS Capstone?

Simplilearn's Data Science Capstone project will give you an opportunity to implement the skills you learned in the Data Scientist Master's Program.



Why DS Capstone?

This Data Science Capstone course end project will validate your expertise as a job -ready data scientist.

You will be able to:

- Apply the knowledge, skills, and capabilities you learned in the Data Science Master's Program
- Build a project on your own from start to finish
- Gain real-world exposure to modern data science challenges
- Learn to build your own models to make better data-driven decisions
- Choose your own domain and technology stack for your project

Project Milestones

The project milestones are as follows:

01

Data Processing - Various data processing techniques are applied to make raw data meaningful.

02

Model Building - Techniques such as regression and decision trees are leveraged to build machine learning models that enable accurate and intelligent predictions.

03

Model Fine -tuning - Various techniques are applied to improve the accuracy of your model and select the champion model that provides the best accuracy.

04

Dashboarding and representing results – Results are required to be exported into a dashboard with meaningful insights using Tableau.

Prerequisites of the Capstone Project

Learner must be proficient in:

Data Science using R, Python, or SAS

and

Data Visualization using Tableau

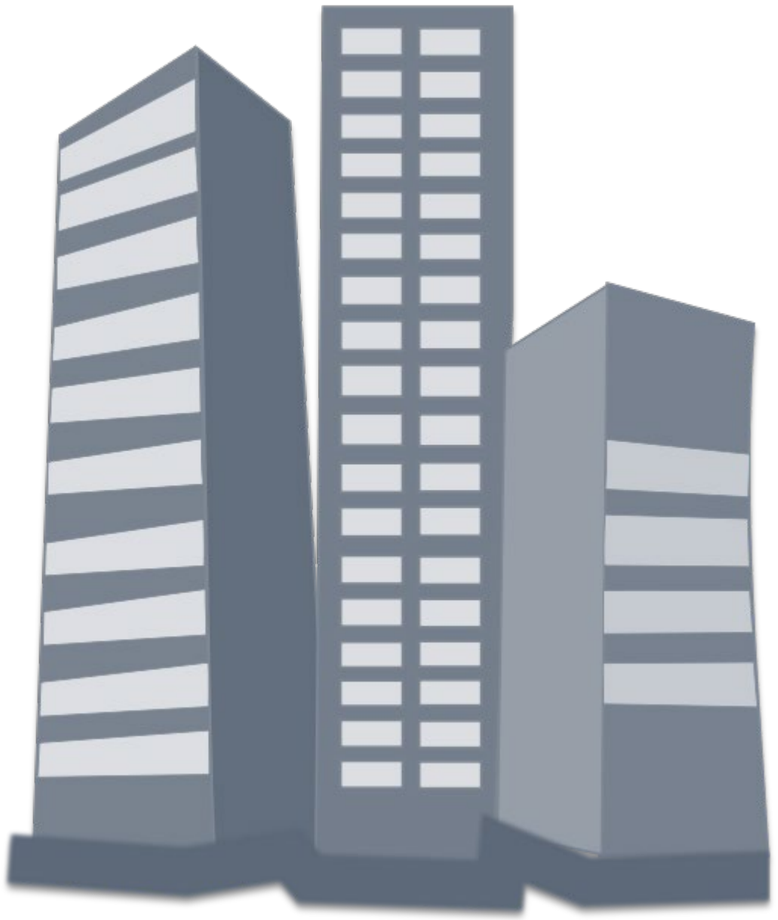
Project 1: Problem Statement

A banking institution requires actionable insights into mortgage-backed securities, geographic business investment, and real estate analysis.

The mortgage bank would like to identify potential monthly mortgage expenses for each region based on monthly family income and rental of the real estate.

A statistical model needs to be created to predict the potential demand for amount of loan in dollars for each of the region in the USA. Also, there is a need to create a dashboard which would refresh periodically, post data retrieval from the agencies.

The dashboard must demonstrate relationships and trends for the key metrics as follows: number of loans, average rental income, monthly mortgage and owner's cost, family income vs mortgage cost comparison across different regions. The metrics described here do not limit the dashboard to these few.



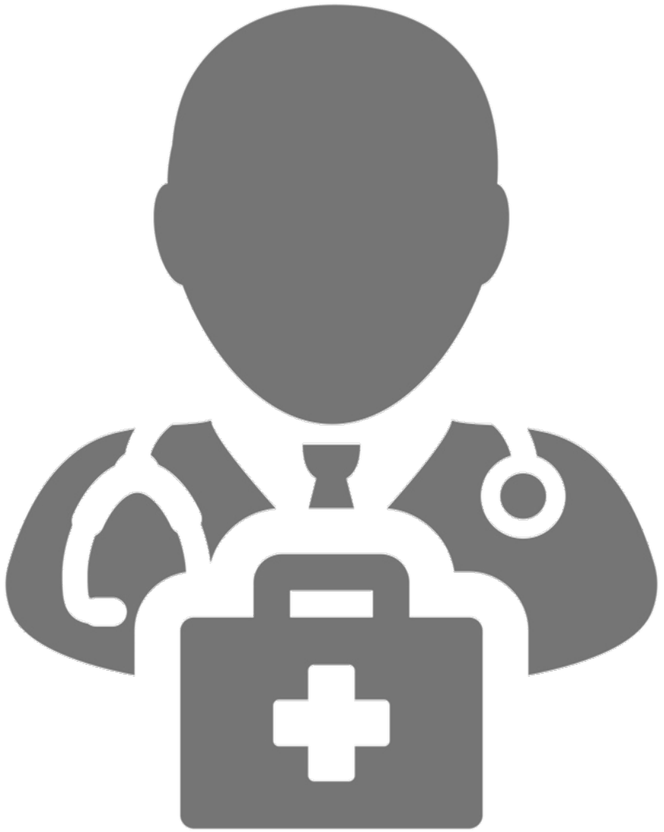
Domain: Real Estate

Project 2: Problem Statement

NIDDK (National Institute of Diabetes and Digestive and Kidney Diseases) research creates knowledge about and treatments for the most chronic, costly, and consequential diseases.

The dataset used in this project is originally from NIDDK. The objective is to predict whether or not a patient has diabetes, based on certain diagnostic measurements included in the dataset.

Build a model to accurately predict whether the patients in the dataset have diabetes or not.



Domain: Healthcare

Project 3: Problem Statement

It is a critical requirement for business to understand the value derived from a customer. RFM is a method used for analyzing customer value.

Customer segmentation is the practice of segregating the customer base into groups of individuals based on some common characteristics such as age, gender, interests , and spending habits

Perform customer segmentation using RFM analysis. The resulting segments can be ordered from most valuable (highest recency, frequency, and value) to least valuable (lowest recency, frequency, and value).

Domain: Retail

Week-wise Agenda

	Project 1 (Real Estate)	Project 2 (Healthcare)	Project 3 (Retail)
Week 1	Data import, data preparation, and data exploration phase 1	Data exploration phase 1	Data wrangling and data transformation
Week 2	Data exploration phase 2	Data exploration phase 2	Data modeling phase 1
Week 3	Data pre-processing	Data modeling phase 1	Data modeling phase 2
Week 4	Data modelling and data reporting	Data modeling phase 2 and data reporting	Data Reporting

Know Your Class

Introduce yourself

Mention your years of experience

State your domain and technology

Thank You