# Week 7 Deliverables

## Team Member's Details

Name: Jessica Gallo

Email: jessica.gallo522@gmail.com

Country: United States of America

Company: Wagner College

Specialization: Data Science

Internship Batch: LISUM12

Date: 09/19/2022

Version: 1.0

Data Intake by: Jessica Gallo

Data storage location:
https://drive.google.com/file/d/1P_oMc6gOBlhw6dY5PxaqxV2swdHMUooK/view

## Problem Description

One of the challenges for all pharmaceutical companies is to understand the persistency of drug as per the physician prescription. To solve this problem ABC pharma company approached an analytics company to automate this process of identification.

## Business Understanding

**ML Problem:**

With an objective to gather insights on the factors that are impacting the persistency, build a classification for the given dataset.

**Target Variable:** Persistency_Flag

**Task:**

- Problem understanding
- Data Understanding
- Data Cleaning and Feature engineering
- Model Development

- Model Selection
- Model Evaluation
- Report the accuracy, precision and recall of both the class of target variable
- Report ROC-AUC as well
- Deploy the model
- Explain the challenges and model selection

## Project Lifecycle Along with Deadline

| Week 7 | 09/19/2022 | Problem description |
|---|---|---|
| Week 8 | 09/26/2022 | Problems in the data |
| Week 9 | 10/02/2022 | Data cleansing and transformation |
| Week 10 | 10/09/2022 | EDA of Data with Final Recommendation |
| Week 11 | 10/16/2022 | EDA Presentation and Proposed Modeling Technique |
| Week 12 | 10/23/2022 | Model Selection and Model Building/Dashboard |
| Week 13 | 10/30/2022 | Final Project Report and Code |

## Data Intake Report

Tabular data details:

| Total number of observations | 3425 |
|---|---|
| Total number of files | 1 |
| Total number of features | 69 |
| Base format of the file | .xlsx |
| Size of the data | 899 KB |

Proposed Approach:

- There are no missing values to clean

## Github Repo Link

https://github.com/Gallo13/DataGlacierProject/tree/main/Week%207