

## Week 9

**Group Name:** Ignotus

### **Team Member Details:**

- Corey Hamren, hamhams86@gmail.com, USA, Self-teaching, Data Science
- Motamen MohammedAhmed, Motamen.salih@hotmail.com, UAE, Zayed University, Data Science
- Francis Kim, fkim39@gmail.com, Fkim39@gmail.com, USA, University of Maryland, Data Science
- Inna Soltsman-Groysman, innasol90@gmail.com, USA, Data Science

**Github Repo link** - [https://github.com/TheHamhams/dg\\_group\\_project](https://github.com/TheHamhams/dg_group_project)

### **Data cleansing and transformation done on the data.**

Corey:

- Lowered dimensionality of 'ntm\_speciality' by aggregating
- Changed binary columns to 0 and 1 by applying a lambda

Motamen:

- Removed ptid variable
- Changed 'persistency\_flag' to binary
- Used Ordinal Encoder on categorical columns
- Applied WOE on categorical columns

Francis:

- Changed binary columns to 0 and 1 values

There are no null values present in the dataset. There are unknown values, but they make up such a large percentage of some columns that they are kept as their own category.

Outliers are present, but not transformed due to lack of domain knowledge. Instead, models that perform better with outliers will be used.

