**Raw Data:**

1. LCDataDictionary.xlsx (Reference for fields/columns)

2. loan\_data\_2007\_2014.csv

**Part 1: Preprocessing:**

(Eg WoEIV file contains ref example for WOE/IV)

(Files required to run loan\_data\_2007\_2014.csv)

Code:

1. Preparation.ipynb

2. Preparation\_test.ipynb (for same preprocessing for test data)

O/p: Data Stored in:

1. loan\_data\_inputs\_test.csv

2. loan\_data\_inputs\_train.csv

3. loan\_data\_targets\_test.csv

4. loan\_data\_targets\_train.csv

**Part 2: PD modelling and Validation:**

(Files required to run: All O/p files from part 1)

Code:

1. PD\_model.ipynb

O/p: Model stored in:

1. pd\_model.sav
2. df\_scorecard.csv
3. inputs \_train\_with\_ref\_cat.csv

**Part 3: LGD, EAD, EL:**

(Files required to run:

1. loan\_data\_2007\_2014\_preprocessed.csv (Preprocessed file from part 1 as required)

2. O/p file pd\_model.sav from part 2

3. O/p files from part 1 with prefix ‘loan\_data\_inputs’)

Code:

1. LGD\_EAD\_EL.ipynb

O/p: Model/Files stored in:

1. lgd\_model\_stage\_1.sav (Stage 1: logistic regression)

2. lgd\_model\_stage\_2.sav (Stage 2: linear regression)

3. loan\_data\_defaults.csv (Data just with default customer entries)

**For more info about data source:** https://www.lendingclub.com/info/statistics.action