The data enclosed in this folder are:

Complaints received by Bureau of Consumer Financial Protection about financial products and services from 2012-2016.

It includes complaint information as follows:

* Date
* Product
* Company
* Complaint description
* State
* Complaint Method
* Response from company and whether it's timely

Using loglinear regressioin, KNN, LDA, you can develop several classification models, depending on your questions, such as:

* which states have more complaints than other states,
* which companies would have more complaints than others?
* which complaint methods would lead more complaints than others?
* what products would have created more complaints

Whilst developing your conclusions, there is a prominent issue at overfitting issue of each model (or under-fitting), so you need to avoid the over (and under-)fitting with cross-validation.

**Question 1. Develop several classification problems based on the data enclosed** (at least 4 problems should be specified )

**Question 2. At each classification problem, propose the best classification model, and explain why it would be** (at each classification problem, you have to apply more than one classification model)**.**

**Question 3. For the best classification model you proposed, i) develop the Python Codes (you can either use my own codes, or sklearn library)**

**Question 4. Now you can compare, for an individual classification problem you have specified, why a classification model is better than the other with such as precision, recall and confusion matrix. Justify your answers in conjunction with question 2, 3 and 4.**

Your answer should be written in Python "Notebook" that I can re-run on my computer.