**Passos iniciais:**

- Importar a base csv;

- Ver o formato das variáveis (describe);

- Transformações de variáveis (número para categóricas) – tanto como fazer esta transformação, como informar o software para ele entender a variável de determinada forma (as\_data, as\_srting).

Por exemplo: Informar que tem variáveis de “data”

- Criar variáveis novas, calculadas a partir de outras variáveis;

* (fat/valor empréstimos);
* fazer o cálculo de prazo (utilizando datas);
* log(variáveis);
* transformar variáveis do tipo categóricas em numérica;

- Selecionar variáveis (separar as quantitativas das categóricas) – criar sub-bases;

- Criar dummies;

- Aprender alguns comandos para análise exploratória dos dados e visualização (hist., box plot, tabela de correlação, tabelas do tipo group by);

- gerar uma sub-base, somente com variáveis tipo.a = PI

- Variáveis categóricas - aprender group by (com base em clientes inadimplentes – qual seria garantia, linha de fin, rating, etc).

- Aprender o If (selecionar variáveis com base numa regra – ex: somente empresas do privado + projetos e investimentos);

- Usar if + group by;

- Variáveis quantitativas – medidas centrais, variabilidade, quartis;

- Testes de correlação (quanti x categórica e categórica x categórica).

**Dúvidas:**

- Como fazer alguma testagem/conferência para ver se os dados coletados estão corretos? Dicas de integridades dos dados;

- Como ir salvando os códigos antigos e alterações nas bases de dados;

- Dicas de quando quer aprender uma técnica nova (comunidades de pyhton, referencias, vídeo-aulas)

- Cronograma (entender a metodologia de desenvolvimento do projeto).

**Técnicas:**

- Quais o prof já usou?

- Indicações de referencias para aprender sobre cada uma das técnicas.

**ANÁLISES QUE PENSO QUE PODEM SER ÚTEIS**

**1 - Modelo de credit rating – Técnicas de classificação:**

- Regressão logística;

- Análise discriminante;

- Random forest;

**Porém, independente da técnica, fazer um pipeline:**

- Para testar modelos com diferentes variáveis e diferentes parâmetros do modelo;

- Validação (k-fold, e cruzada);

- Avaliação.

**2 - Fazer fator com variáveis:**

- Análise fatorial (aprimoramento de modelos) – Modelos dentro de modelos – Análise do componente principal

**3 - Agrupar clientes com características semelhantes – Técnicas de Agrupamento:**

- Conhecer os métodos de calcular a distância;

- Agrupamento;

- Selecionando o melhor k (técnica do joelho)

**4- Relação provisão com fatores de risco:**

- Regressão linear (aplicando regularização ridge e lasso);