1. Dataset: "ml-20ml" from MovieLens
2. Software: Python, Jupyter notebook
3. Package:
   a. Initial analysis
   ```
   import pandas as pd
   import numpy as np
   import matplotlib.pylab as plt
   %matplotlib inline
   import seaborn as sns
   ```
   b. Memory-based model
   ```
   from sklearn.neighbors import NearestNeighbors
   from sklearn.model_selection import train_test_split
   from sklearn.metrics import mean_squared_error, mean_absolute_error
   import numpy as np
   import pandas as pd
   from scipy.sparse import csr_matrix
   import matplotlib.pyplot as plt
   import warnings
   from pandarallel import pandarallel
   ```
   c. Model-based model
   ```
   import os
   import pyspark
   from pyspark import SparkFiles
   from pyspark import SparkContext
   from pyspark.sql import SQLContext
   from pyspark.sql.functions import rand, col
   from pyspark.ml.recommendation import ALS
   from pyspark.ml.evaluation import RegressionEvaluator
   from pyspark.ml.tuning import ParamGridBuilder, TrainValidationSplit
   ```