```mermaid
flowchart TD

  Import[Import data for inspiration]

  Pose[Pose statistical question: movie rating distributions across Adventure, Comedy and Drama genres are the same]

  Wrangle[Data wrangling: omit data rows with missing values, change date format]

  EDA[Exploratory data analysis: explore the relevant data columns visually]

  Import --> Pose --> Wrangle --> EDA
```

Simulation and statistical tests function

INPUTS:

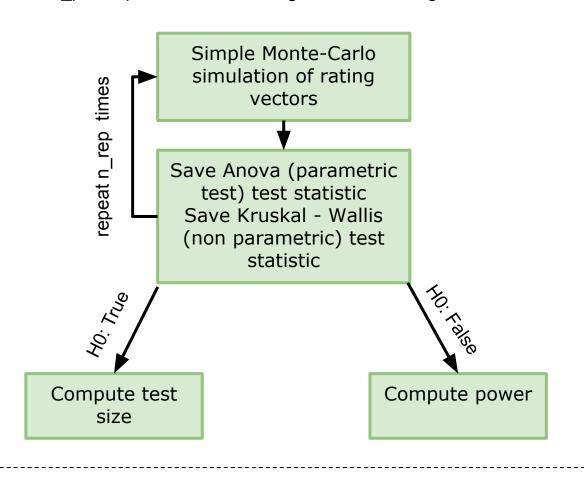   n - sample size,

    n_rep - times to compute power,

    H0 - null hypothesis, whether TRUE or FALSE,

    ad_prob - probabilities of ratings for adventure genre,

    co_prob - probabilities of ratings for comedy genre,

    dr_prob - probabilities of ratings for adventure genre.

- Simple Monte-Carlo simulation of rating vectors
- Save Anova (parametric test) test statistic Save Kruskal - Wallis (non parametric) test statistic

repeat n_rep times

H0: True → Compute test size

H0: False → Compute power

Scenario 1 (H0 = FALSE): simulate data samples similar to the movielens data

Scenario 2 (H0 = FALSE): increase sample size to get 90% power

Scenario 3 (H0 = TRUE): change theoretical effect size to 0, desirable test size = 5%

Scenario 4 (H0 = FALSE): change theoretical effect size to very small, desirable power = 90%