Netflix Business case study

Analyze the data and generate insights that could help Netflix decide which type of shows/movies to produce and how to grow the business.

Major Steps:

- Data Preprocessing
  - Solve nested data challenge for the following columns: Director, Cast, Listed In, Country
  - Treat Duration Column
  - Date Time Columns
  - Rating Column
- Data Cleaning
  - Multiple Level Imputation
- EDA
  - Univariate
  - Bivariate

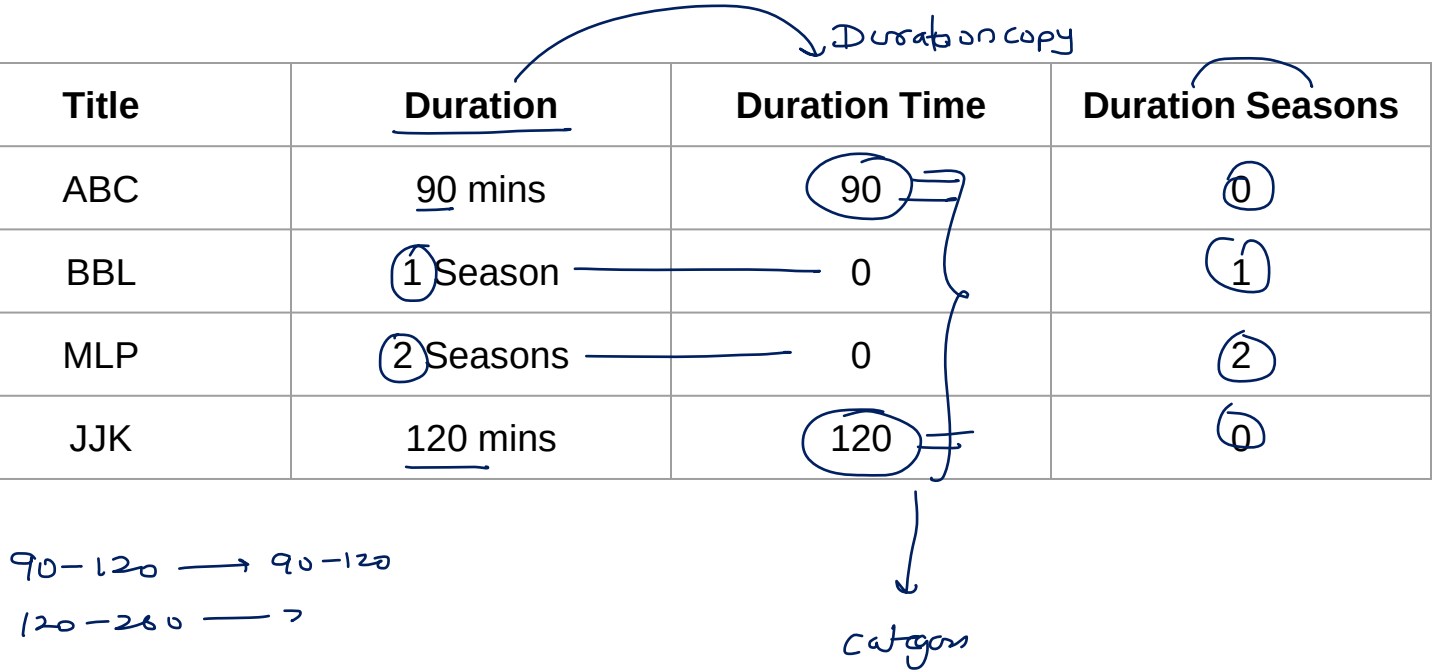
Netflix Business case study

Movie Title	Actor
Scaler-A Data Scientist	Anshuman, Abhimanyu
Jhola	Ankit, Bhavish
Myjus	Ravee, Divya
Hamara Bajaj	Bajaj

Movie Title	Actor
Scaler-A Data Scientist	Anshuman
Scaler-A Data Scientist	Abhimanyu
Jhola	Ankit
Jhola	Bhavish
Myjus	Ravee
Myjus	Divya
Hamara Bajaj	Bajaj

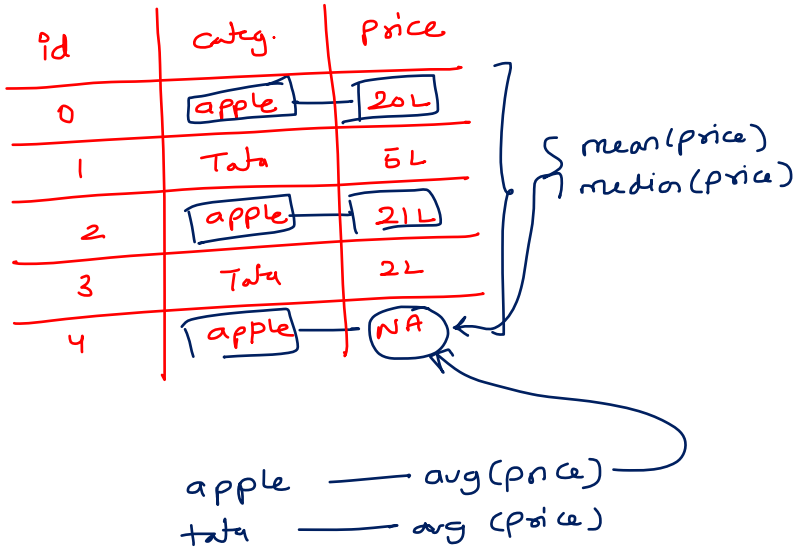
Netflix Business case study

Title	Duration	Duration Time	Duration Seasons
ABC	90 mins	90	0
BBL	1 Season	0	1
MLP	2 Seasons	0	2
JJK	120 mins	120	0



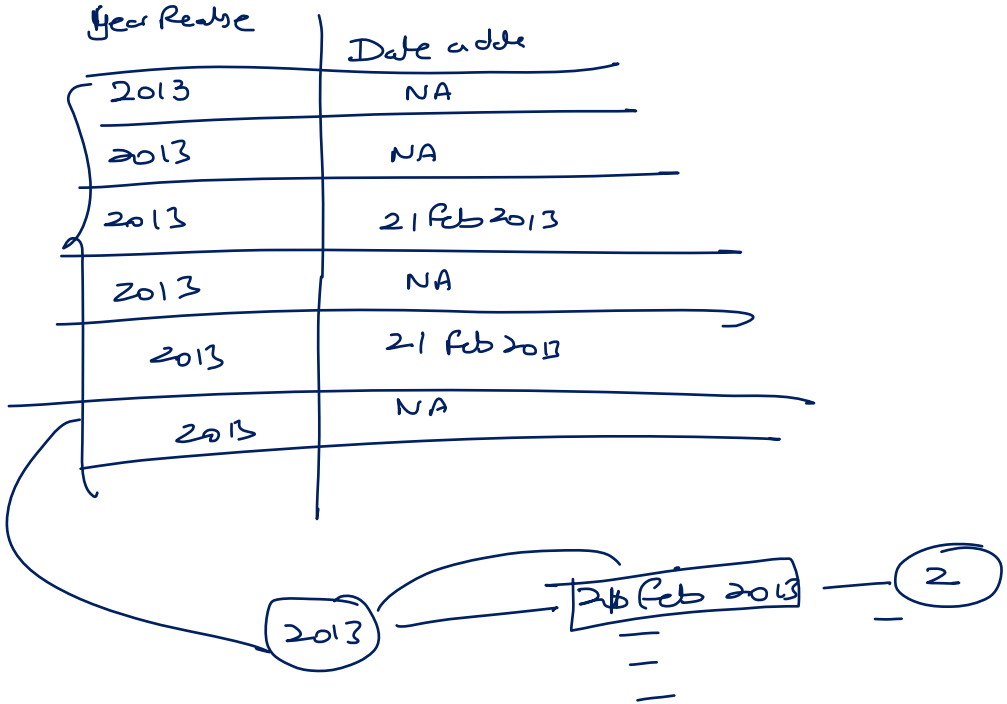
Netflix Business case study

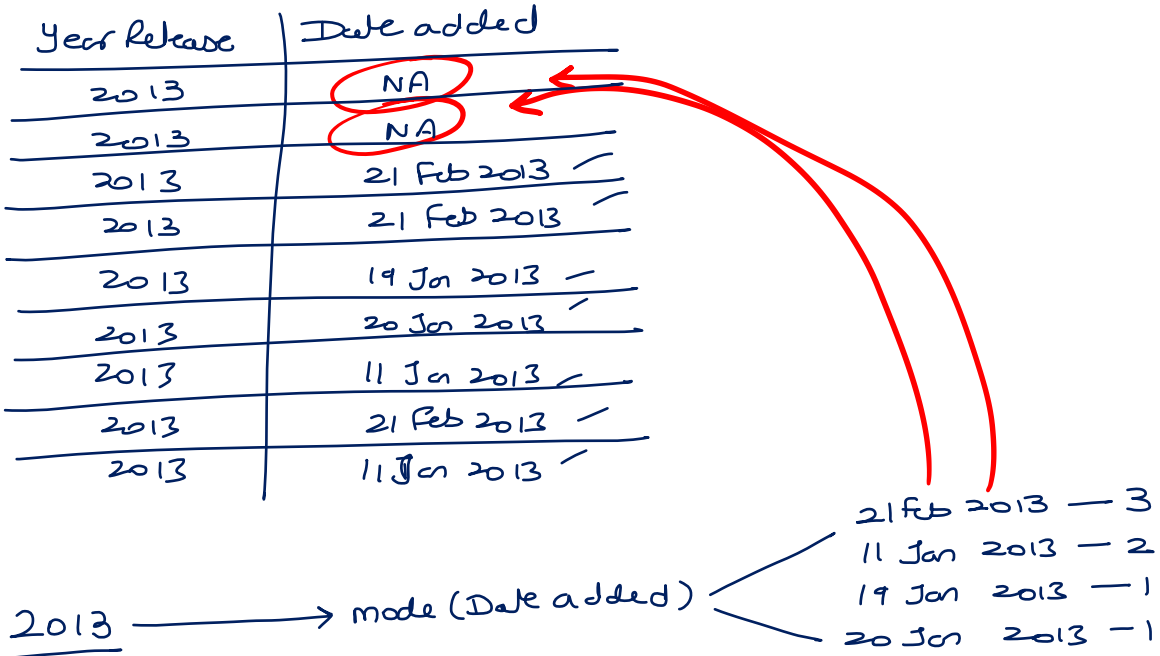
Date Added	Month Added	Week Day Added	Year Added
September 25, 2021	9	Saturday	2021
September 24, 2021	9	Friday	2021
September 29, 2021	9	Wednesday	2021



800+
2M rows
6 months
=

Category imputation


Year Release
Date added
2013
NA
2013
NA
2013
21 Feb 2013
2013
NA
2013
21 Feb 2013
2013
NA
2013
2013
24 Feb 2013
2

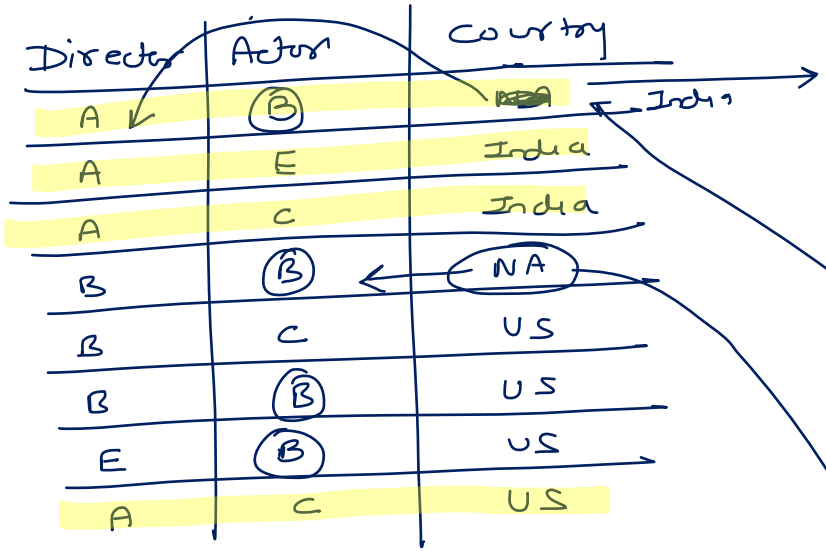

Year Release
Date added
2013
NA
2013
NA
2013
21 Feb 2013
2013
21 Feb 2013
2013
19 Jan 2013
2013
20 Jan 2013
2013
11 Jan 2013
2013
21 Feb 2013
2013
11 Jan 2013
21 Feb 2013 — 3
11 Jan 2013 — 2
19 Jan 2013 — 1
20 Jan 2013 — 1
2013 → mode (Date added)
2015 — mode (Date added)
2016 — mode (Date
2019

Dateadded	yearRelease
NA	2013
NA	2013
NA	2013
NA	2014
NA	2014
NA	2015
NA	2015
NA	2019

unique(yearRelease) ⟶ [2013, 2014, 2015, 2019]

=

for i in [2013, 2014, 2015, 2019]:



A → NA
A → India
A → India
A → US

A → mode (country) → India

B → [redacted] India
B → NA
B → US
B → US

B → mode (country) = US