

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

The diagram illustrates the transformation of the 'Duration' column into two new columns: 'Duration Time' and 'Duration Seasons'. A curved arrow labeled 'Duration copy' points from the 'Duration' column to the 'Duration Time' column. Another curved arrow points from the 'Duration Time' column to the 'Duration Seasons' column. The 'Duration' column contains values like '90 mins', '1 Season', '2 Seasons', and '120 mins'. The 'Duration Time' column contains numerical values (90, 0, 0, 120) which are circled. The 'Duration Seasons' column contains numerical values (0, 1, 2, 0) which are also circled. Handwritten lines connect '1 Season' to '0' and '2 Seasons' to '0' in the 'Duration Time' column. A bracket on the right side of the 'Duration Time' column groups the values 90, 0, 0, and 120. Below the table, there are handwritten notes: '90-120 -> 90-120' and '120-260 -> ?' on the left, and 'categories' with an arrow pointing down from the bracketed values in the 'Duration Time' column on the right.

Title	<u>Duration</u>	Duration Time	Duration Seasons
ABC	<u>90</u> mins	90	0
BBL	1 Season	0	1
MLP	2 Seasons	0	2
JJK	<u>120</u> mins	120	0

90-120 → 90-120

120-260 → ?

categories

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

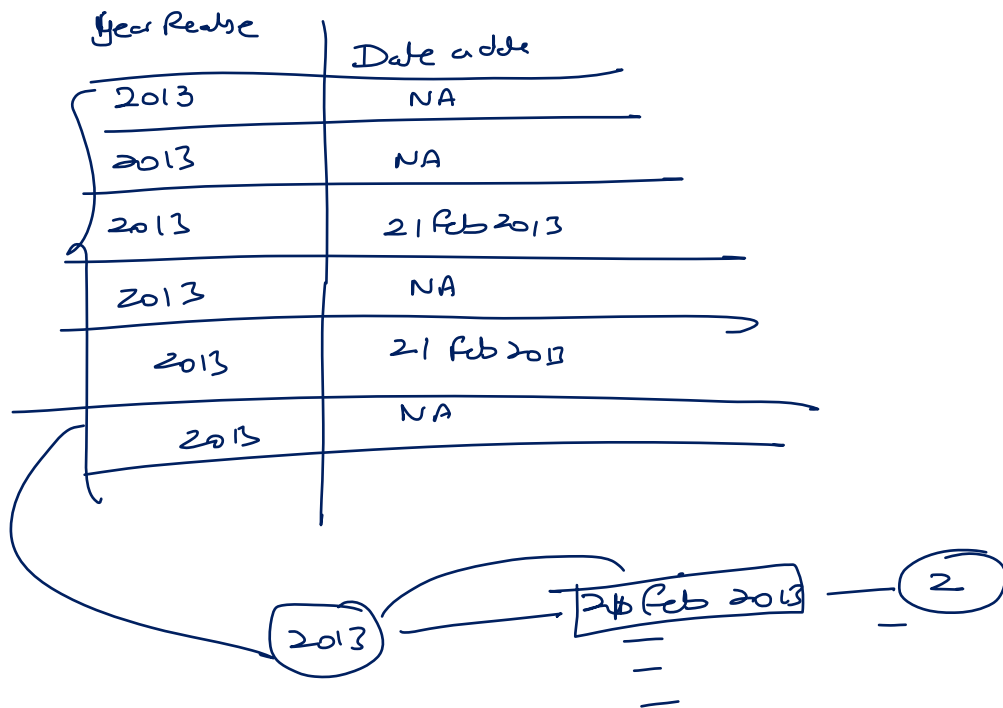
800+
2m rows
5 months
=

id	categ.	price
0	apple	20L
1	Tata	5L
2	apple	21L
3	Tata	2L
4	apple	NA

mean(price)
median(price)

apple — avg(price)
tata — avg(price)

Category imputation



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 ✓

2013 → mode (Date added)

- 21 Feb 2013 — 3
- 11 Jan 2013 — 2
- 19 Jan 2013 — 1
- 20 Jan 2013 — 1

2015 — mode (Date added)

2016 — mode (Date)

2019

Date added	Year Release
NA	2013
NA	2013
NA	2013
NA	2014
NA	2014
NA	2015
NA	2015
NA	2019

Unique (year Release) \longrightarrow [2013, 2014, 2015, 2019]
=

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

Director	Actor	Country
A	(B)	NA India
A	E	India
A	C	India
B	(B)	NA
B	C	US
B	(B)	US
E	(B)	US
A	C	US

A → NA
 A → India
 A → India
 A → US

A → mode(country) → India

B → ~~NA~~ India
 B → NA
 B → US
 B → US

B → mode(country) = US