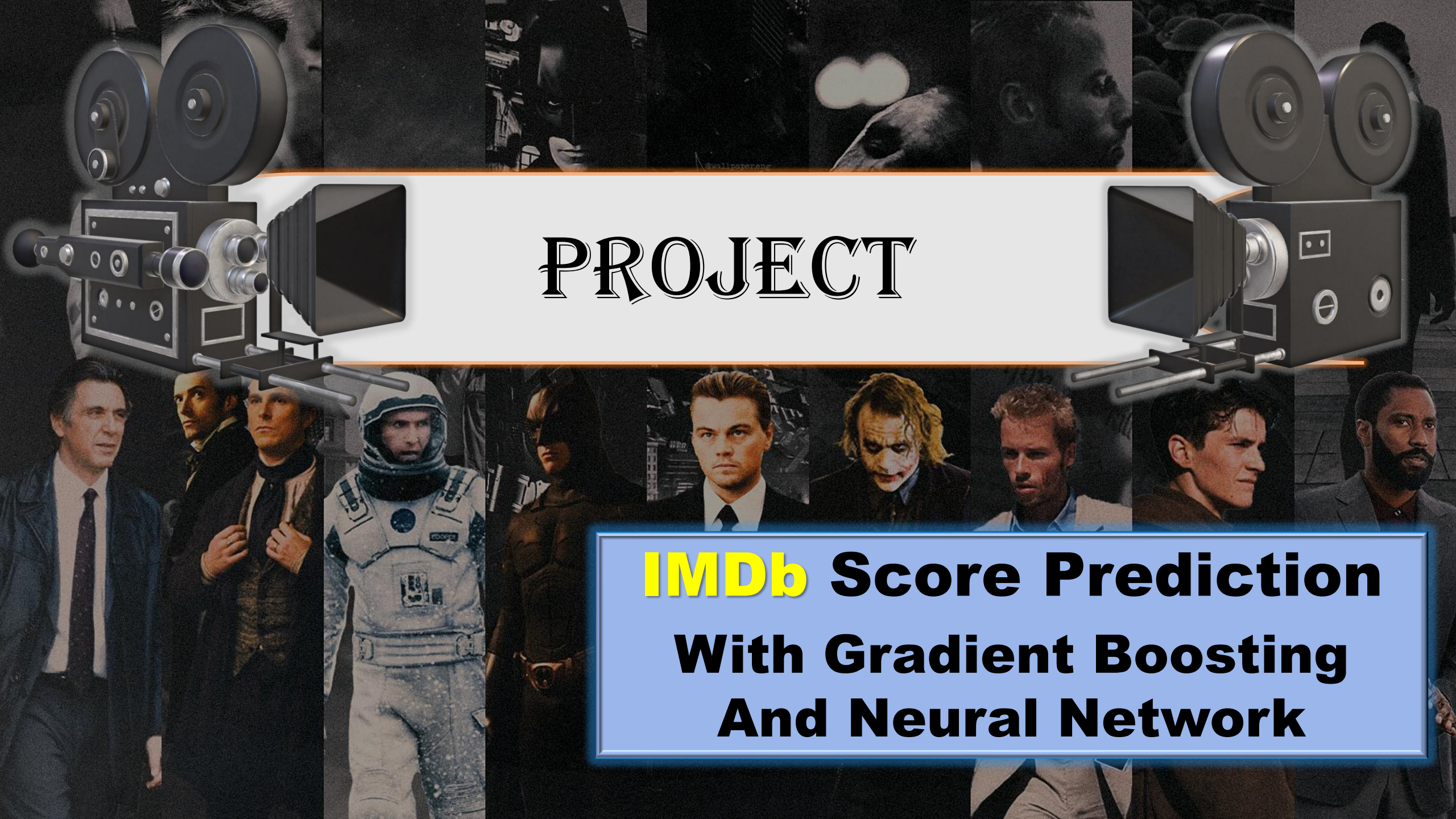

PROJECT
IMDb Score Prediction
With Gradient Boosting
And Neural Network

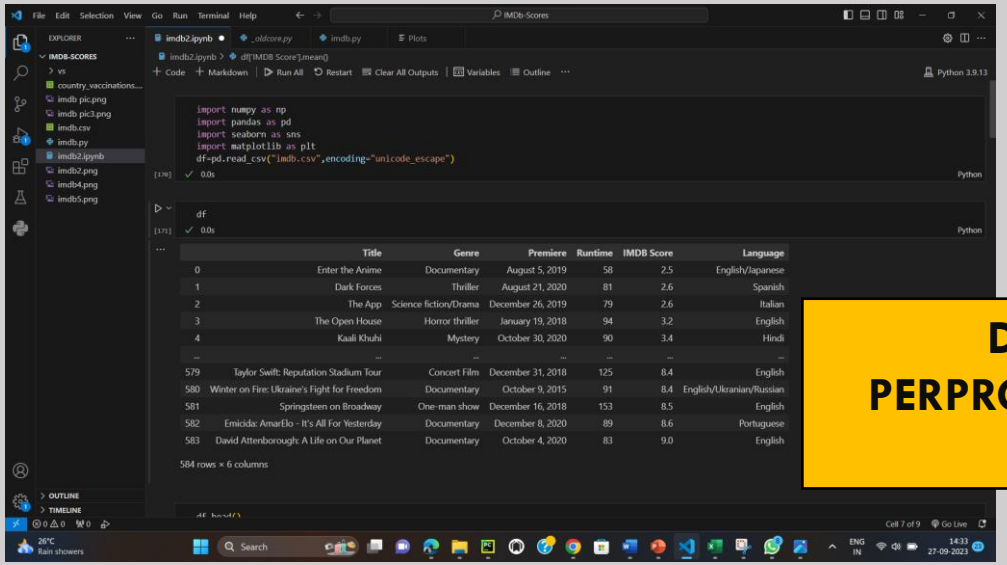
Develop a machine learning model to predict the IMDb Scores of the movies available on Films Based in their genre, premiere date, runtime and language . The model aims to accurately finds the popularity of the movies to the assist users in discovering highly rated films that align with their preferences

Design the project based on :

* Data Source
* Data Preprocessing
* Feature Engineering
* Model Selection
* Model Training
* Evaluation

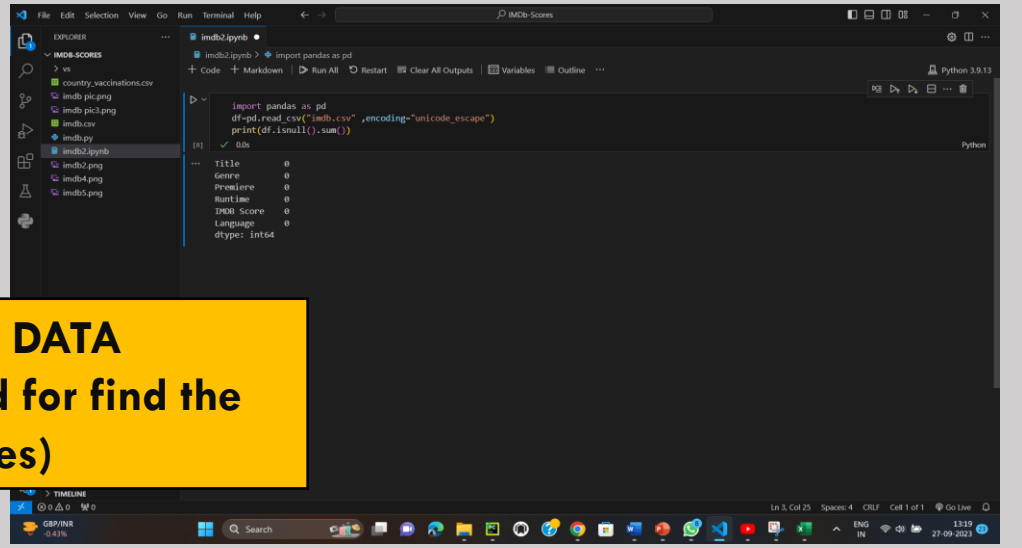
Main algorithms & ML are :

linear Regression ,Random Forest Regression to Predict IMDb Scores .Train the selected model using preprocessing data Regression metrics like MAE ,MSE,R-squared,Gradient Boosting and Neural Network.

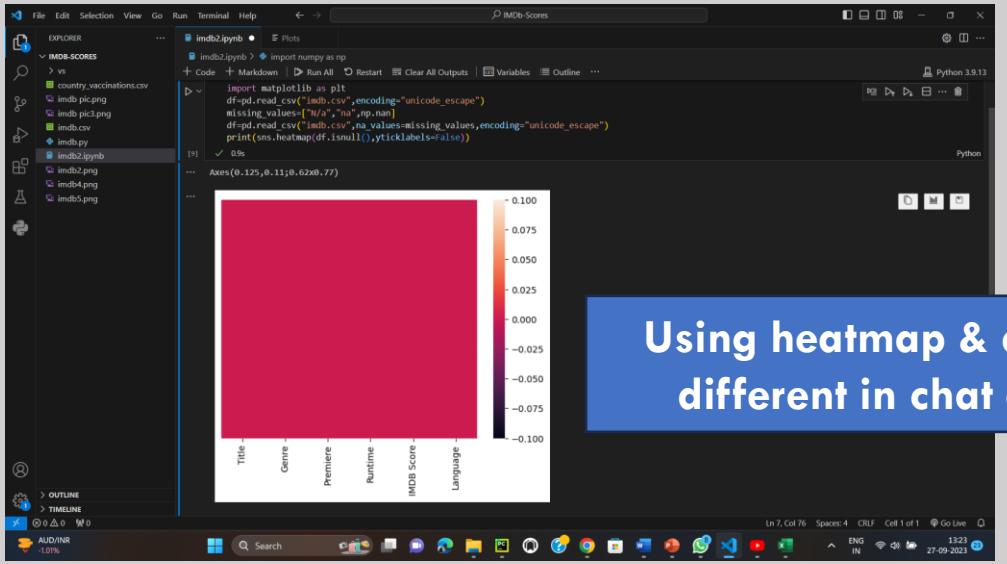


	Title	Genre	Premiere	Runtime	IMDB Score	Language
0	Enter the Anime	Documentary	August 5, 2019	58	2.5	English/Japanese
1	Dark Forces	Thriller	August 21, 2020	81	2.6	Spanish
2	The App	Science fiction/Drama	December 26, 2019	79	2.6	Italian
3	The Open House	Horror thriller	January 19, 2018	94	3.2	English
4	Kaali Khuhi	Mystery	October 30, 2020	90	3.4	Hindi
...	...	...	...	...	...	...
579	Taylor Swift: Reputation Stadium Tour	Concert Film	December 31, 2018	125	8.4	English
580	Winter on Fire: Ukraine's Fight for Freedom	Documentary	October 9, 2015	91	8.4	English/Ukranian/Russian
581	Springsteen on Broadway	One-man show	December 16, 2018	153	8.5	English
582	Emicida: AmarElo - It's All For Yesterday	Documentary	December 8, 2020	89	8.6	Portuguese
583	David Attenborough: A Life on Our Planet	Documentary	October 4, 2020	83	9.0	English

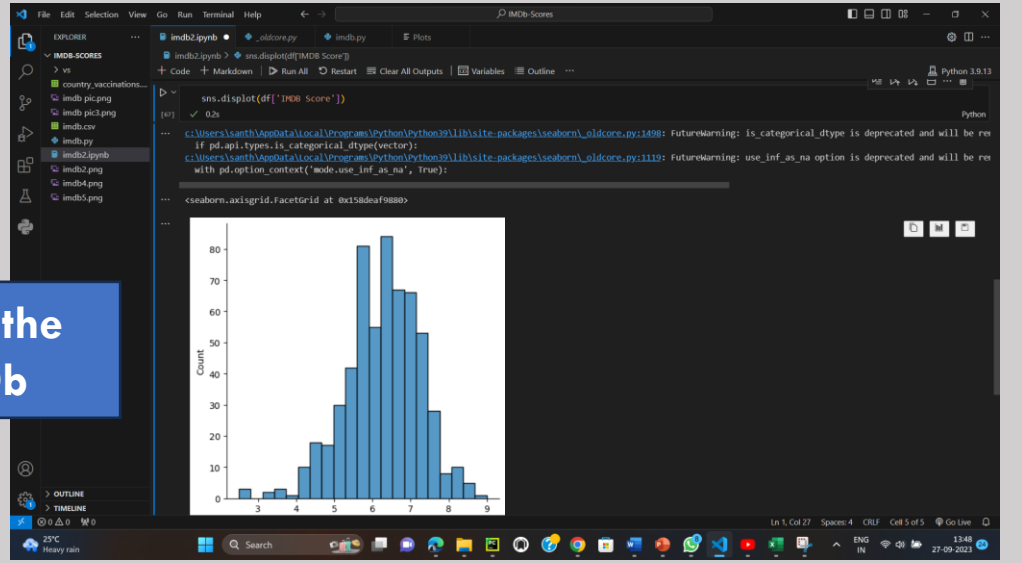
584 rows × 6 columns



DATA SOURCE & DATA PERPROCESSING (used for find the missing values)

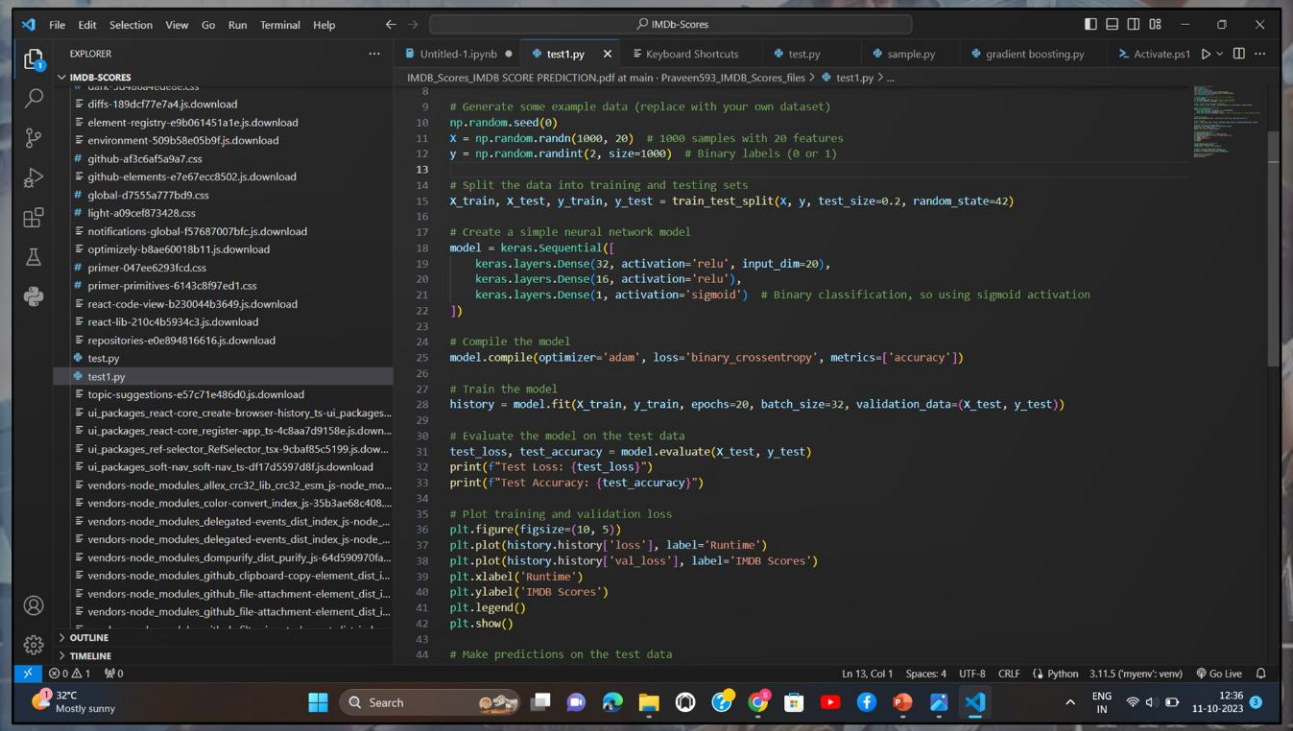


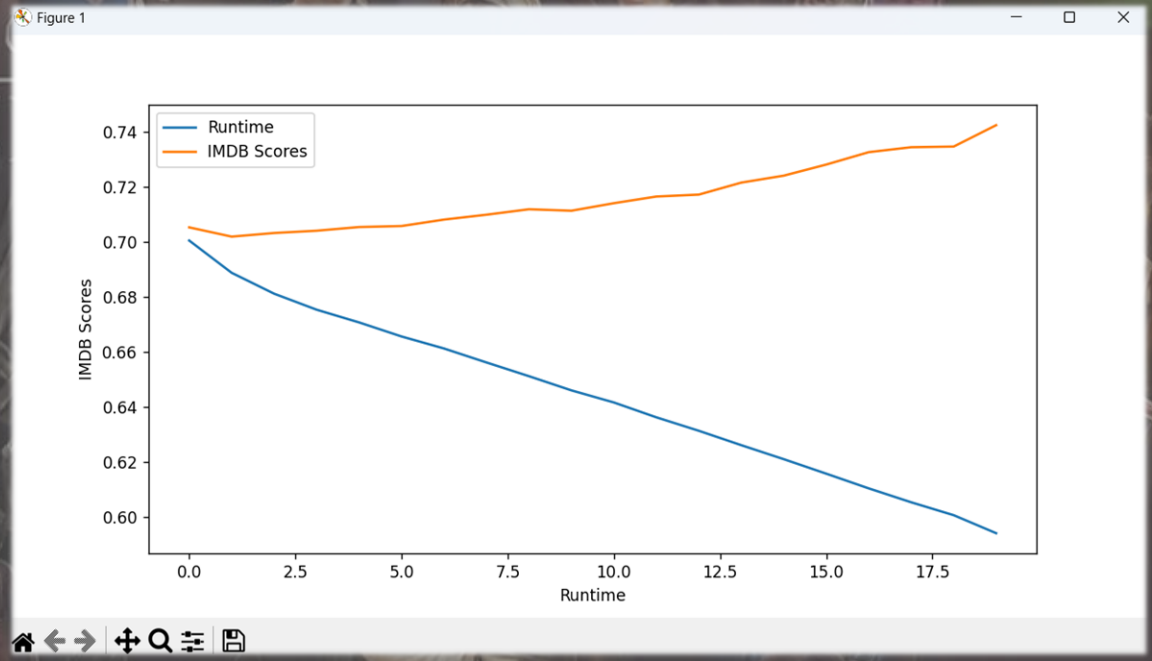
Using heatmap & displot can find the different in chat and predict IMDb



NEURAL NETWORKING

- Neural Networking is an ensemble learning technique that combines the predictions of several weak learners to create a strong predictive model.
- Neural Networking,with its iterative learning approach, stands as a powerful tool for predictive modeling, providing accurate results across diverse dataset





- **Hyperparameter Tuning:** Proper tuning of hyperparameters is crucial for optimal performance.
- **Potential for Overfitting:** Care must be taken to avoid overfitting, especially if the weak learners are too complex.

Key Components For Gradient Boosting

- **Weak Learners**: Typically shallow decision trees are used.
- **Boosting**: Models are built sequentially, and each subsequent model corrects errors made by the previous ones.

Use Cases For Gradient Boosting

- **Commonly Applied**: Used in various domains, including finance, healthcare, and Kaggle competitions.

Key Components For Neural Network

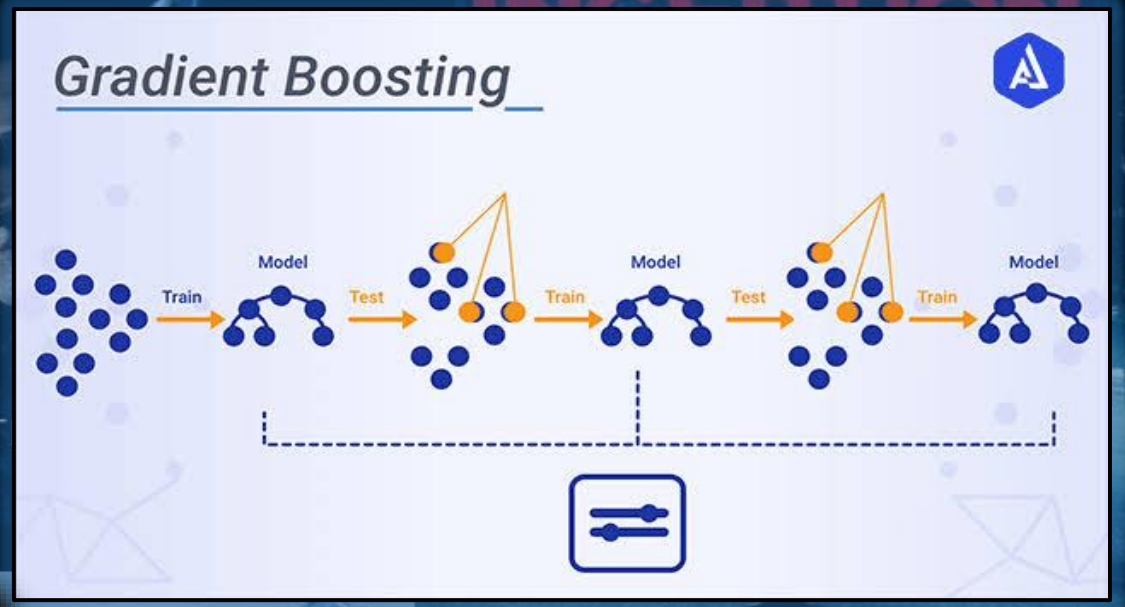
- **Neurons**: Basic computational units that process information.
- **Layers**: Organized in input, hidden, and output layers.
- **Activation Functions**: Determine the strength of connections between neurons.

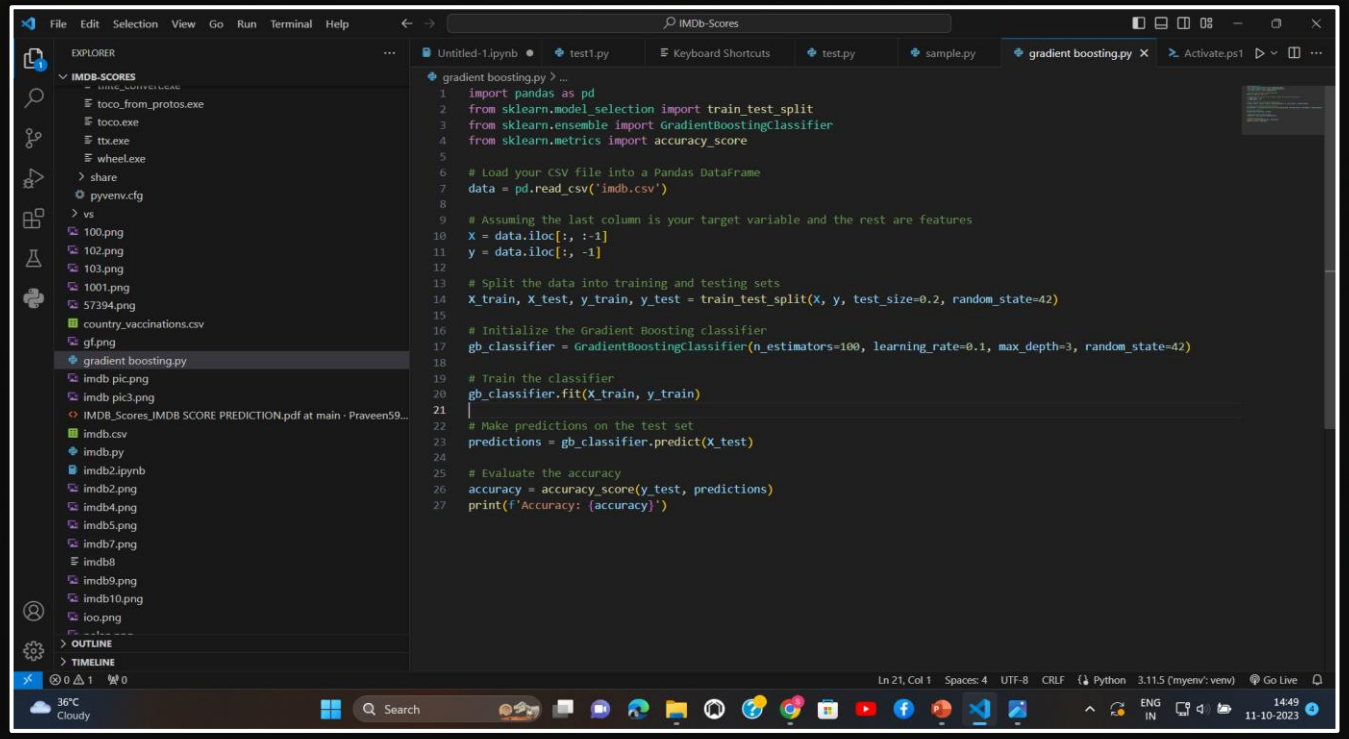
Use Cases For Gradient Boosting

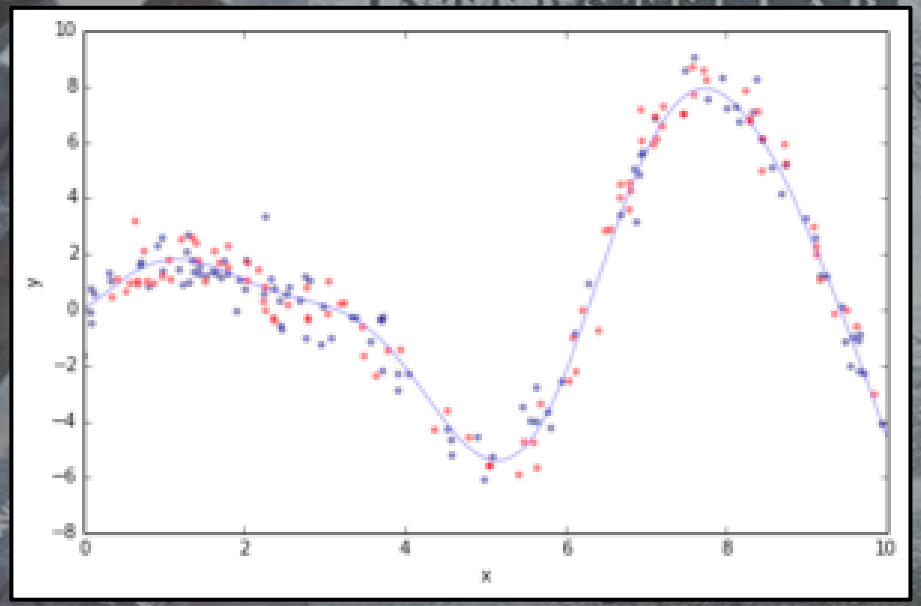
- **Image and Speech Recognition**: Neural Networks excel in tasks like image and speech recognition.

GRADIENT BOOSTING

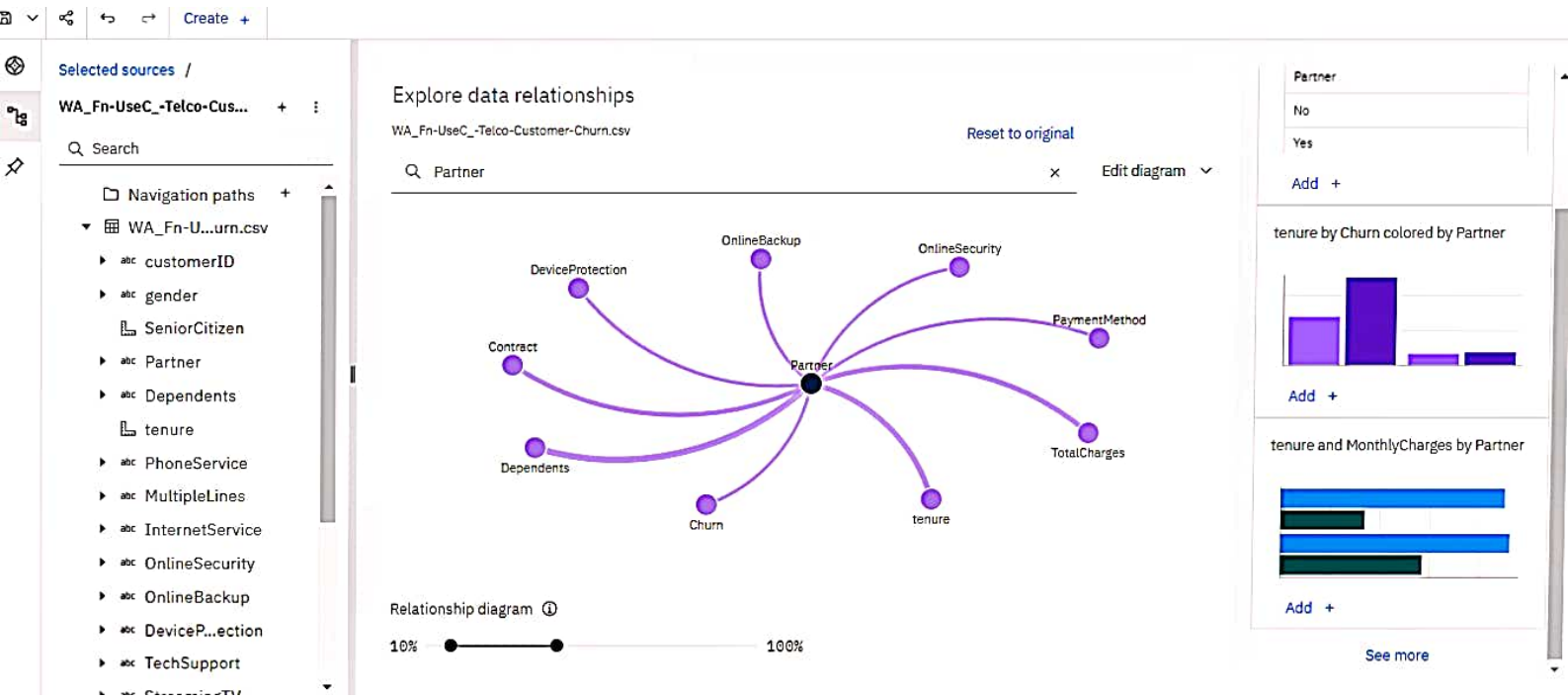
- Gradient Boosting is an ensemble learning technique that combines the predictions of several weak learners to create a strong predictive model.
- Gradient Boosting, with its iterative learning approach, stands as a powerful tool for predictive modeling, providing accurate results across diverse datasets.







THANK YOU


Create +
Selected sources /
WA_Fn-UseC_-Telco-Cus...
Search
Navigation paths
WA_Fn-U...urn.csv
customerID
gender
SeniorCitizen
Partner
Dependents
tenure
PhoneService
MultipleLines
InternetService
OnlineSecurity
OnlineBackup
DeviceP...ection
TechSupport
Explore data relationships
WA_Fn-UseC_-Telco-Customer-Churn.csv
Reset to original
Partner
Edit diagram
OnlineBackup
DeviceProtection
OnlineSecurity
PaymentMethod
Contract
Partner
TotalCharges
Dependents
Churn
tenure
Relationship diagram
10%
100%
Partner
No
Yes
Add +
tenure by Churn colored by Partner
Add +
tenure and MonthlyCharges by Partner
Add +
See more

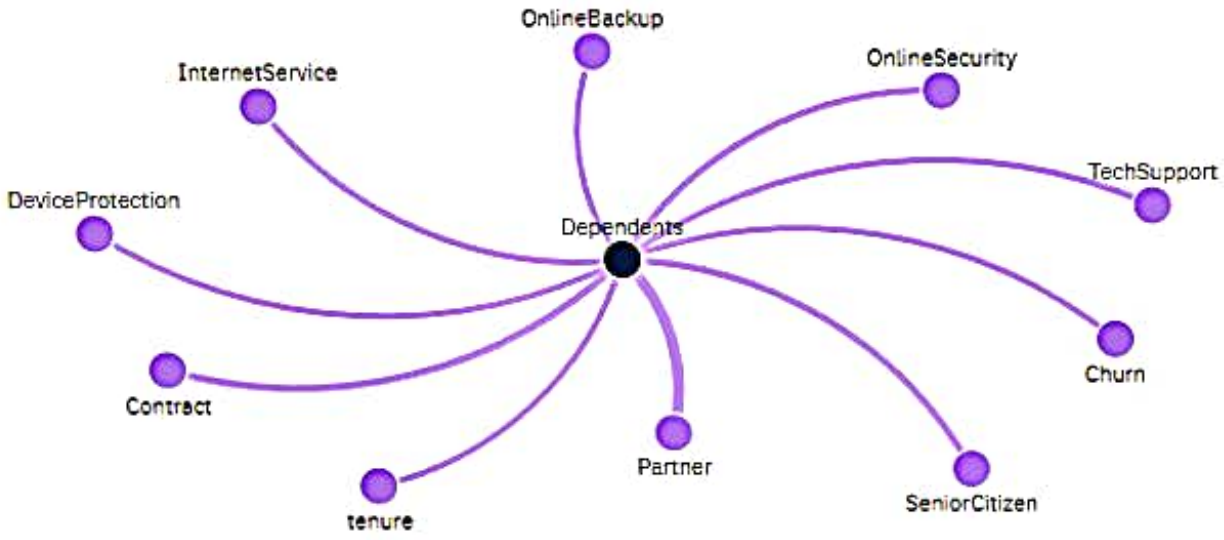
Explore data relationships

WA_Fn-UseC_-Telco-Customer-Churn.csv

Reset to original

Dependents

Edit diagram



Relationship diagram

10% 100%

Cards

tenure by Chur... by customerID

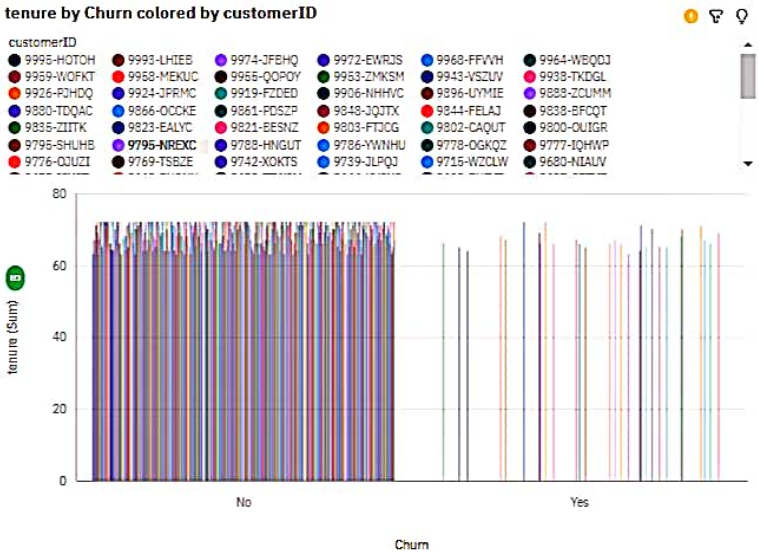
1

MonthlyCharges...red by gender

2

tenure and Mon...ges by Partner

Data relationships



Details

Over all values of **Churn** and **customerID**, the sum of **tenure** is nearly 87 thousand.

The summed values of **tenure** range from 63 to 72.

For **tenure**, the most significant values of **customerID** are 8809-XKHMD, 8204-YJCLA, 2274-XUATA, 0244-LGNFY, and 9919-FZDED, whose respective **tenure** values add up to 360, or 0.4 % of the total.

For **tenure**, the most significant value of **Churn** is No, whose respective **tenure** values add up to over 81 thousand, or 93.8 % of the total.

For **tenure**, the most significant values of **customerID** are 8809-XKHMD, 0244-LGNFY, 2274-XUATA, 8204-YJCLA, and 9919-FZDED, whose respective **tenure** values add up to 360, or 0.4 % of the total.

Create + 2/9 Column Related Compare Analytics Details Fields Properties

Cards

tenure by Chur... by customerID

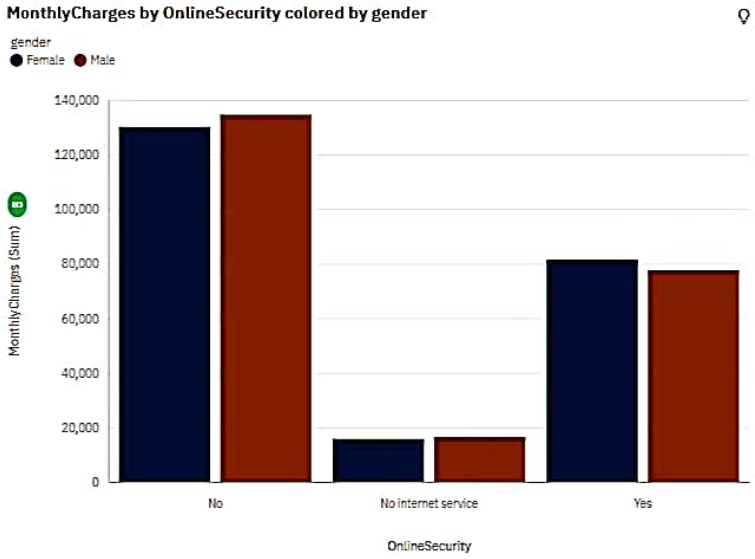
1

MonthlyCharges...red by gender

2

tenure and Mon...ges by Partner

Data relationships



Details

MonthlyCharges is unusually low when **OnlineSecurity** is No internet service.

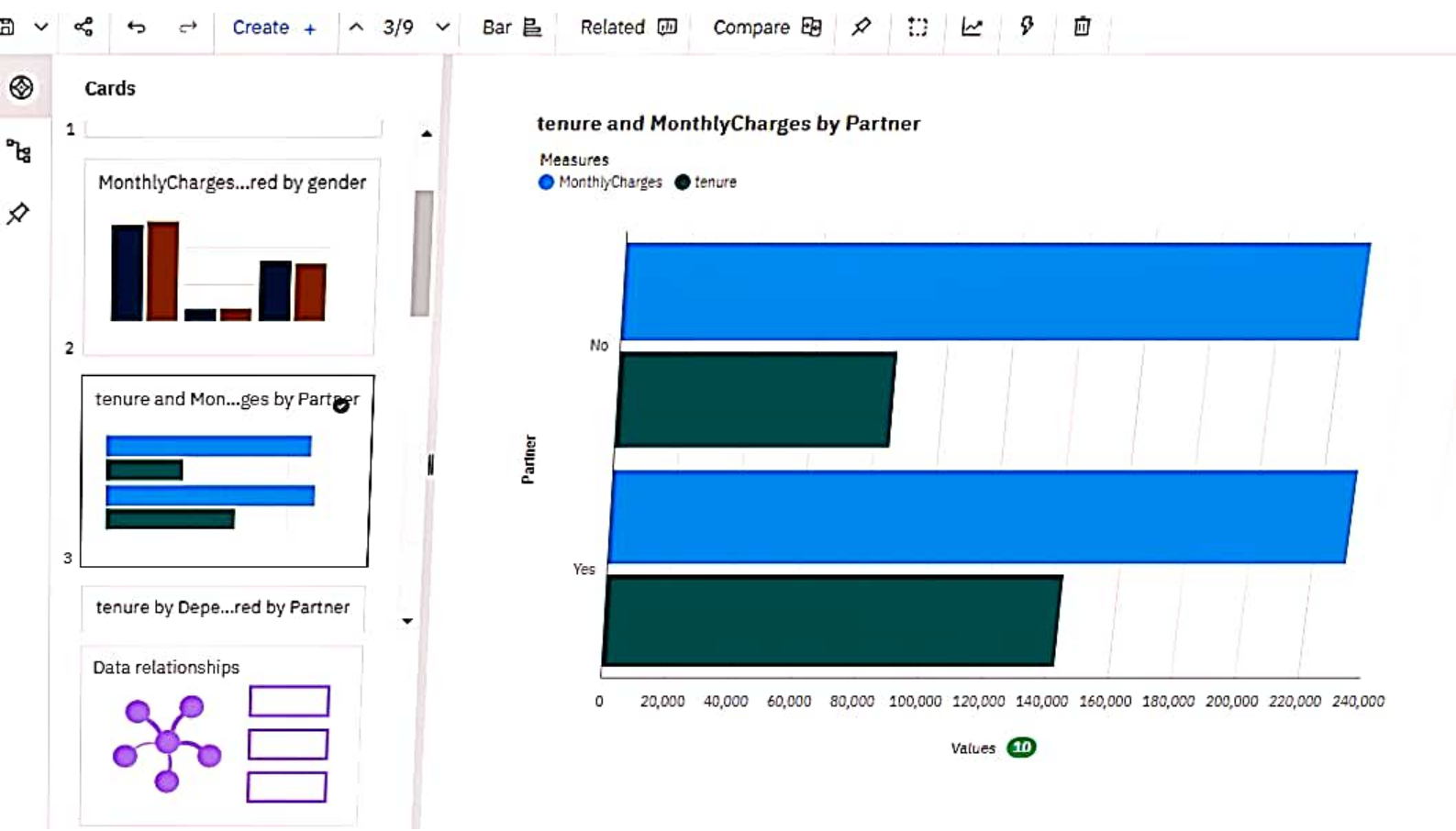
Across all values of **OnlineSecurity** and **gender**, the sum of **MonthlyCharges** is over 456 thousand.

The summed values of **MonthlyCharges** range from nearly 16 thousand to almost 135 thousand.

MonthlyCharges is unusually low when the combinations of **OnlineSecurity** and **gender** are No internet service and Female and No internet service and Male.

For **MonthlyCharges**, the most significant values of **OnlineSecurity** are No and Yes, whose respective **MonthlyCharges** values add up to nearly 424 thousand, or 92.9 % of the total.

For **MonthlyCharges**, the most significant value of **gender** is Male, whose respective


Create
3/9
Bar
Related
Compare
Cards
1
MonthlyCharges...red by gender
2
tenure and Mon...ges by Partner
3
tenure by Depe...red by Partner
Data relationships
tenure and MonthlyCharges by Partner
Measures
MonthlyCharges
tenure
Partner
No
Yes
0
20,000
40,000
60,000
80,000
100,000
120,000
140,000
160,000
180,000
200,000
220,000
240,000
Values
10

Cards

2

tenure and Mon...ges by Partner

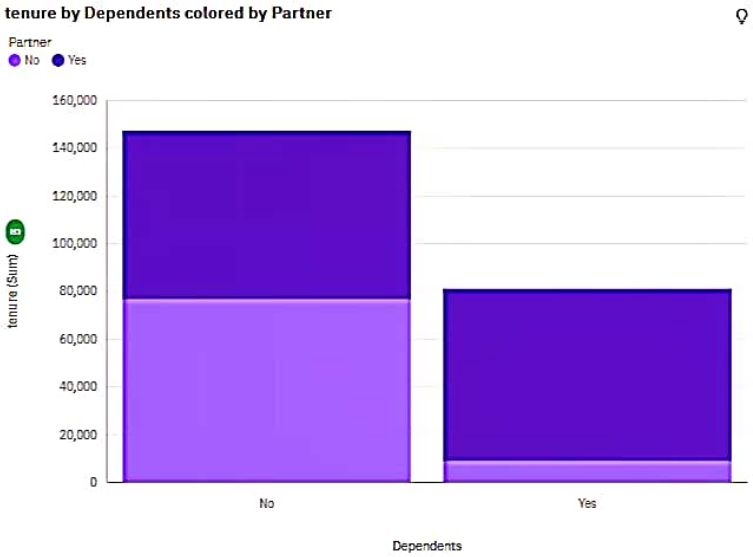
3

tenure by Depe...red by Partner

4

MonthlyCharges... PhoneService

Data relationships



Details

Over all values of **Dependents** and **Partner**, the sum of **tenure** is almost 228 thousand.

The summed values of **tenure** range from over 8500 to over 76 thousand.

tenure is unusually low when the combination of **Dependents** and **Partner** is Yes and No.

For **tenure**, the most significant value of **Dependents** is No, whose respective **tenure** values add up to over 147 thousand, or 64.5 % of the total.

For **tenure**, the most significant value of **Partner** is Yes, whose respective **tenure** values add up to almost 143 thousand, or 62.7 % of the total.

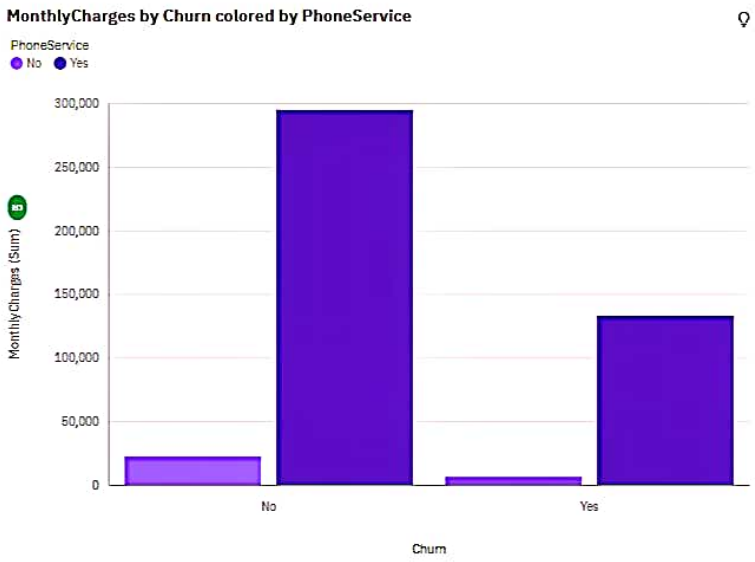
Cards

4

5 MonthlyCharges... PhoneService

TotalCharges by... MultipleLines

Data relationships



Details

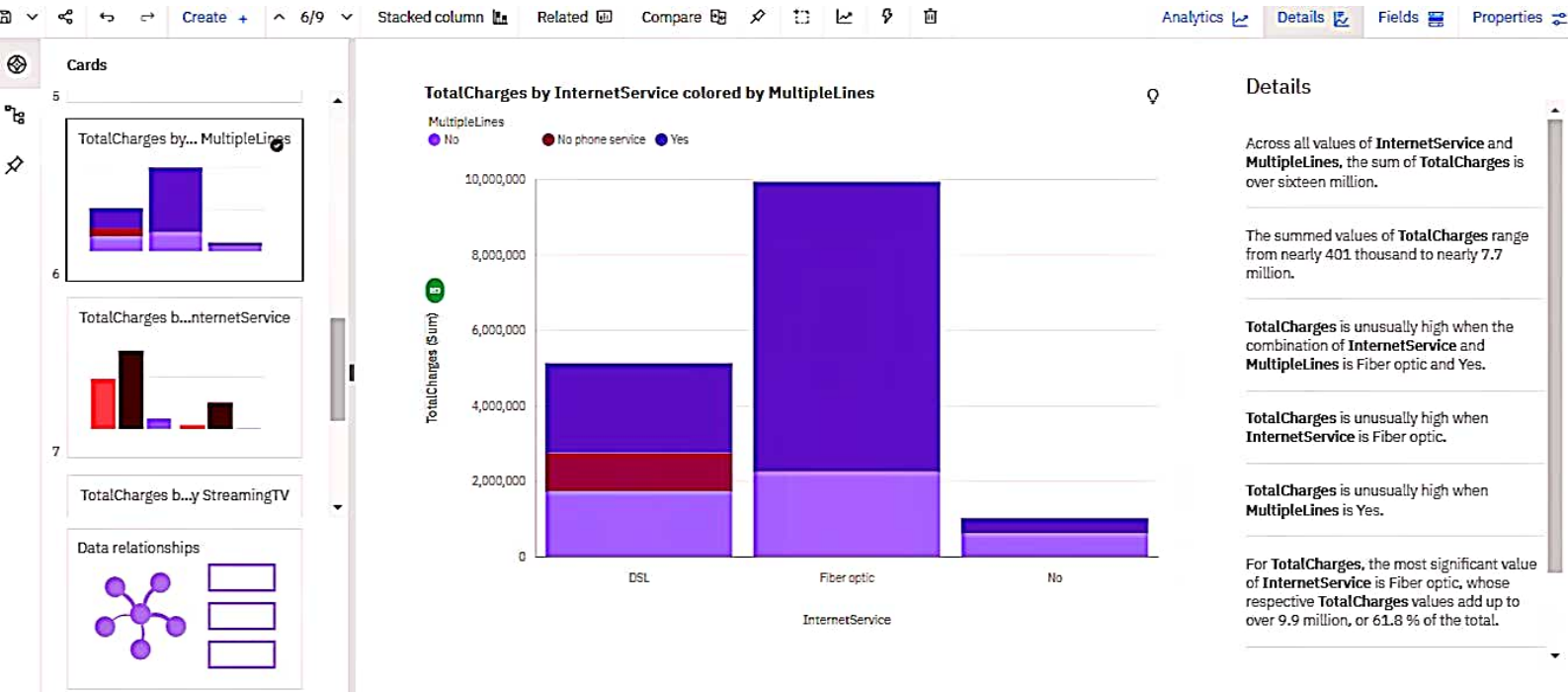
Across all values of **Churn** and **PhoneService**, the sum of **MonthlyCharges** is over 456 thousand.

The summed values of **MonthlyCharges** range from nearly 6500 to nearly 295 thousand.

MonthlyCharges is unusually high when the combination of **Churn** and **PhoneService** is No and Yes.

For **MonthlyCharges**, the most significant value of **PhoneService** is Yes, whose respective **MonthlyCharges** values add up to over 427 thousand, or 93.7 % of the total.

For **MonthlyCharges**, the most significant value of **Churn** is No, whose respective **MonthlyCharges** values add up to nearly 317 thousand, or 69.5 % of the total.


Create
6/9
Stacked column
Related
Compare
Analytics
Details
Fields
Properties
Cards
TotalCharges by... MultipleLines
TotalCharges b...nternetService
TotalCharges b...y StreamingTV
Data relationships
TotalCharges by InternetService colored by MultipleLines
MultipleLines
No
No phone service
Yes
10,000,000
8,000,000
6,000,000
4,000,000
2,000,000
0
TotalCharges (Sum)
DSL
Fiber optic
No
InternetService
Details
Across all values of InternetService and MultipleLines, the sum of TotalCharges is over sixteen million.
The summed values of TotalCharges range from nearly 401 thousand to nearly 7.7 million.
TotalCharges is unusually high when the combination of InternetService and MultipleLines is Fiber optic and Yes.
TotalCharges is unusually high when InternetService is Fiber optic.
TotalCharges is unusually high when MultipleLines is Yes.
For TotalCharges, the most significant value of InternetService is Fiber optic, whose respective TotalCharges values add up to over 9.9 million, or 61.8 % of the total.

Cards

5

TotalCharges by... MultipleLines

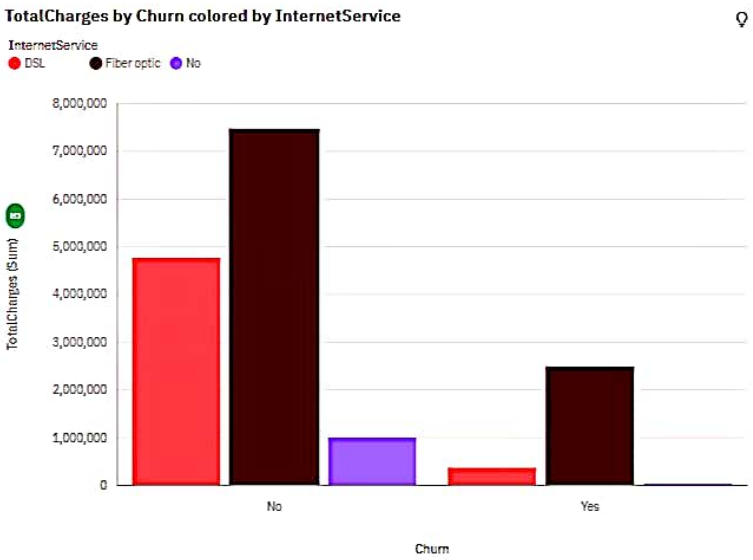
6

TotalCharges b...nternetService

7

TotalCharges b...y StreamingTV

Data relationships



Details

TotalCharges is unusually high when **InternetService** is Fiber optic.

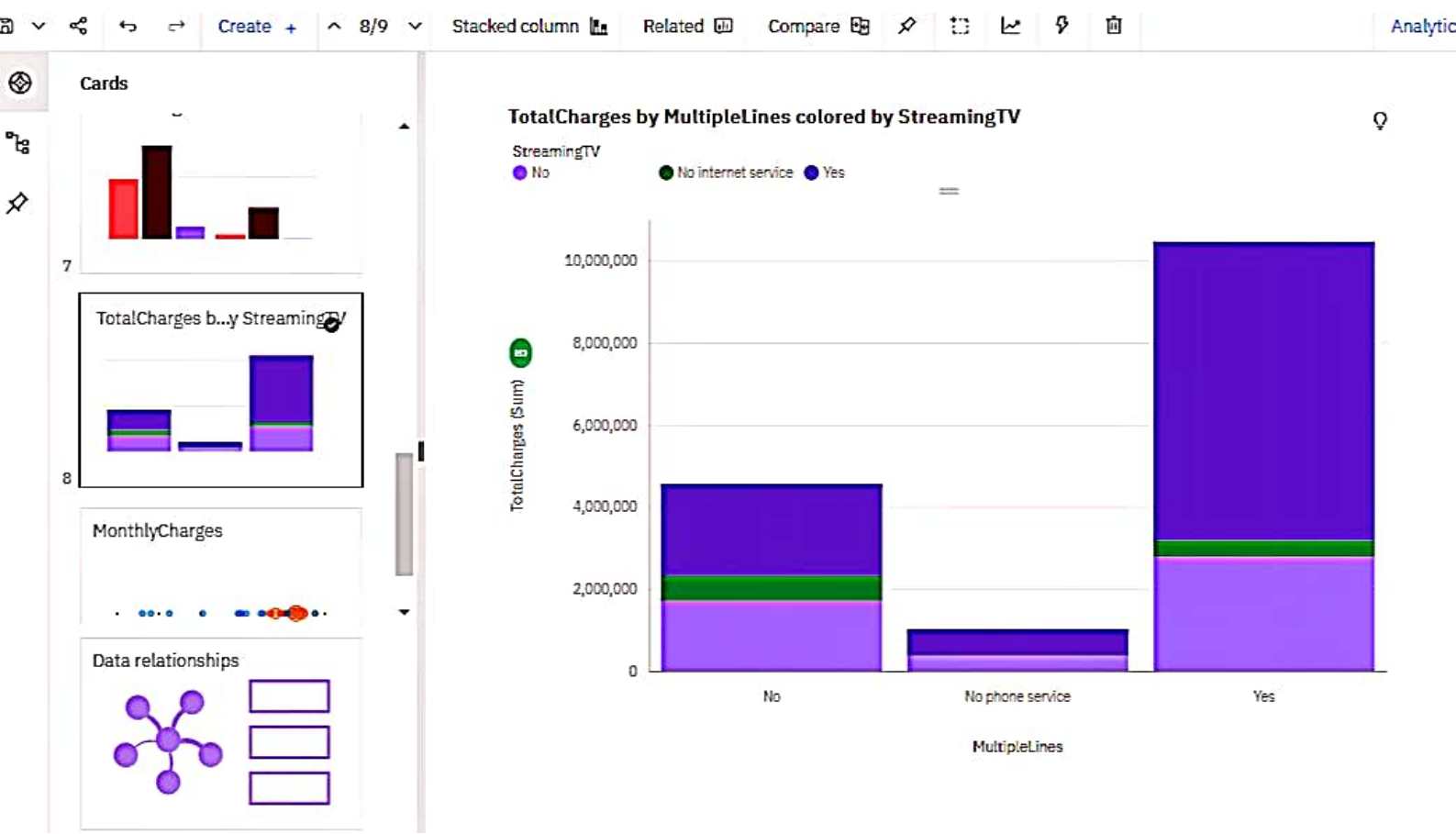
Across all values of **Churn** and **InternetService**, the sum of **TotalCharges** is over sixteen million.

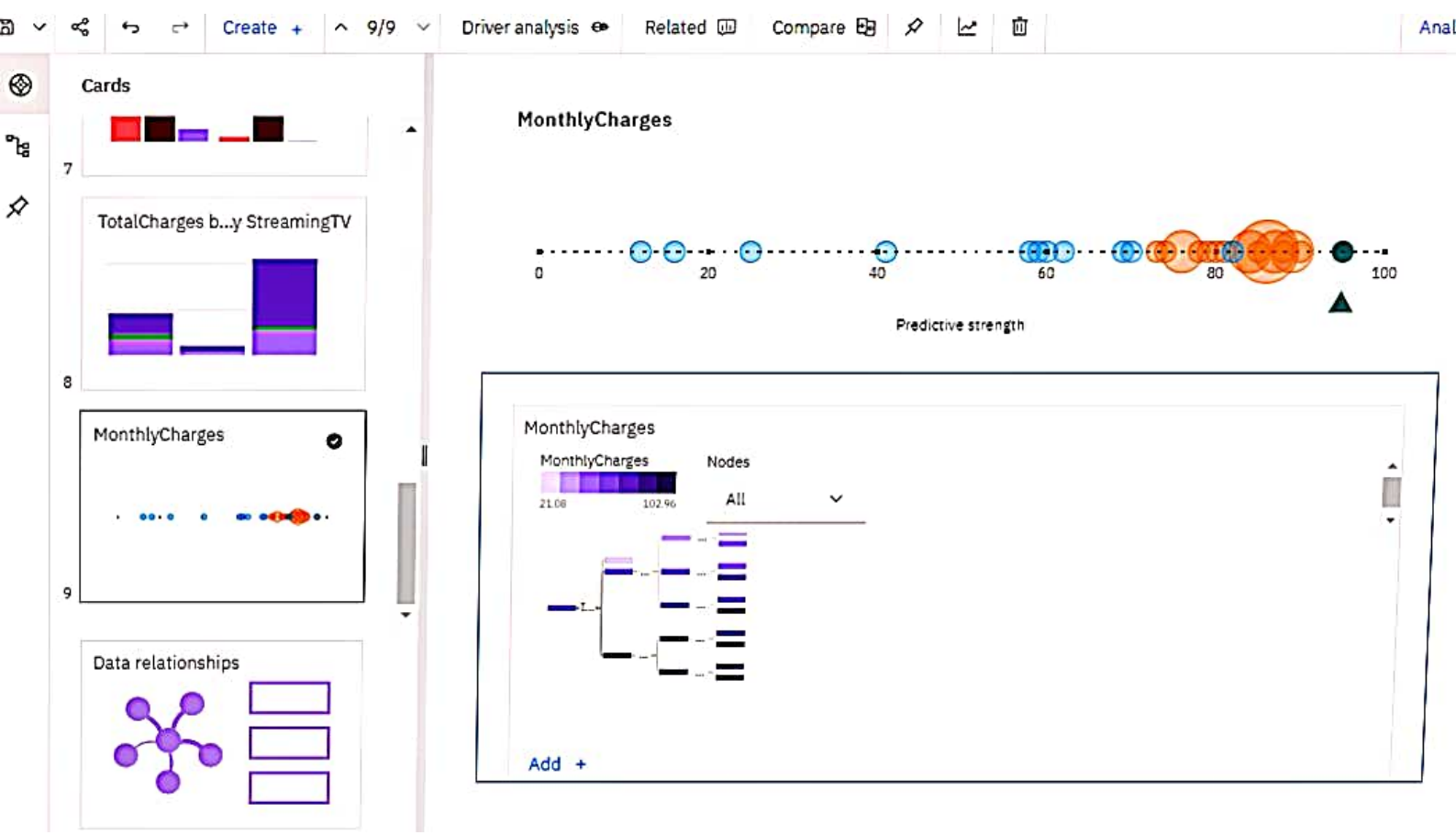
The summed values of **TotalCharges** range from nearly 20 thousand to over 7.4 million.

TotalCharges is unusually high when the combination of **Churn** and **InternetService** is No and Fiber optic.

For **TotalCharges**, the most significant value of **Churn** is No, whose respective **TotalCharges** values add up to over thirteen million, or 82.2 % of the total.

For **TotalCharges**, the most significant value of **InternetService** is Fiber optic, whose respective **TotalCharges** values add up to over 9.9 million, or 61.8 % of the total.


Create
8/9
Stacked column
Related
Compare
Analytic
Cards
7
TotalCharges b...y StreamingTV
8
MonthlyCharges
Data relationships
TotalCharges by MultipleLines colored by StreamingTV
StreamingTV
No
No internet service
Yes
10,000,000
8,000,000
6,000,000
4,000,000
2,000,000
0
TotalCharges (Sum)
No
No phone service
Yes
MultipleLines


Create +
9/9
Driver analysis
Related
Compare
Cards
7
TotalCharges b...y StreamingTV
8
MonthlyCharges
9
Data relationships
MonthlyCharges
0
20
40
60
80
100
Predictive strength
MonthlyCharges
MonthlyCharges
21.08
102.96
Nodes
All
Add +