Problem Definition:

The problem we aim to address is the need for an improved movie and TV show recommendation system. Existing platforms like IMDb provide basic recommendations based on user ratings and genre preferences. However, these recommendations often lack personalization and depth, leading to user dissatisfaction and missed opportunities for discovering content they would enjoy.

Design Thinking Approach:

1. *Empathize:* Understand the Users

   - Conduct user interviews, surveys, and data analysis to gain insights into what users find lacking in current recommendation systems.

   - Identify common pain points, such as generic recommendations, limited diversity in content suggestions, and difficulty in discovering hidden gems.

2. *Define:* Problem Statement

   - Define the problem statement: "Create a movie and TV show recommendation system that offers personalized, diverse, and engaging content suggestions to users based on their unique preferences and viewing history."

3. *Ideate:* Generate Solutions

   - Brainstorm various features and technologies that can address the defined problem.

   - Consider incorporating machine learning algorithms, user profiling, and content analysis to improve recommendation accuracy.

4. *Prototype:* Create a Mockup

   - Develop a prototype of the IMDb-like platform with new features and recommendations.

   - Include options for users to rate, review, and provide feedback on movies and shows to enhance personalization.

5. *Test:* Gather Feedback

   - Launch the prototype to a select group of users for testing.

   - Collect feedback on the effectiveness of the new recommendation system and user satisfaction.

   - Make necessary adjustments based on user input.

6. *Iterate:* Refine and Enhance

   - Continuously refine the recommendation algorithms based on user interactions and feedback.

   - Explore incorporating natural language processing (NLP) to analyze user reviews and provide better recommendations.

   - Optimize the user interface for ease of use and engagement.

7. *Implement:* Develop the Full System

- Build the full IMDb-like platform with the refined recommendation system and user-friendly interface.

- Implement robust security measures to protect user data and privacy.

- Scale the system to handle a growing user base.

8. *Evaluate:* Measure Success

- Continuously track user engagement, content ratings, and user feedback.

- Use metrics such as click-through rates, user retention, and user-generated content to assess the platform's success.

- Make ongoing improvements based on data-driven insights.

9. *Scale:* Expand and Adapt

- Consider expanding the platform to support other languages and regions.

- Adapt to changing user preferences and technological advancements.

- Explore partnerships with content providers and streaming services to enhance the content library.

By following this design thinking approach, we can create an IMDb-like platform that not only addresses the existing problems but also continuously adapts and improves to meet the evolving needs and preferences of its users.

Introduction:

Building upon the foundation established in Phase 1, Phase 2 focuses on innovating the IMDb score prediction process. The goal is to create a cutting-edge and highly accurate IMDb score prediction model that enhances the recommendation system. Here's the approach:

1. Data Enrichment:

- Expand the dataset by including additional information such as user demographics, viewing history, and content details (e.g., director, cast, keywords).

- Incorporate real-time user interaction data, including watch history, user preferences, and user-generated reviews.

2. Advanced Machine Learning Techniques:

- Implement state-of-the-art machine learning algorithms, such as deep learning models, for IMDb score prediction.

- Use recurrent neural networks (RNNs) or transformer models to capture sequential user behavior and preferences over time.

- Explore techniques like reinforcement learning to optimize user engagement and recommendation accuracy.

3. Feature Engineering:

- Create new features that consider factors like user sentiment in reviews, content popularity trends, and content release dates.

- Leverage natural language processing (NLP) to extract sentiment and insights from user reviews and integrate this information into the prediction model.

4. Explainable AI:

- Develop an explainable AI model that can provide users with clear explanations for IMDb score predictions.

- Allow users to understand why a particular movie or show is recommended and how it relates to their viewing history and preferences.

5. User-Generated Content:

- Encourage users to contribute more detailed reviews, ratings, and comments.

- Utilize user-generated content for sentiment analysis and content understanding, which can further enhance IMDb score predictions.

6. Feedback Loop:

- Implement a robust feedback loop that allows users to rate IMDb score predictions.

- Use this feedback to fine-tune the prediction model in real-time, ensuring that it adapts to changing user preferences and evolving content trends.

7. Ethical Considerations:

- Ensure the responsible use of AI in predictions to avoid biases and ethical issues.

- Regularly audit the algorithms to minimize discrimination and promote fairness in recommendations and IMDb score predictions.

8. Integration with Content Providers:

- Collaborate with content providers and streaming services to access real-time content data.

- Integrate this data to offer users the latest and most relevant recommendations.

9. Continuous Monitoring and Improvement:

- Monitor the accuracy and user satisfaction of IMDb score predictions.

- Implement a continuous improvement process, including A/B testing of prediction models to fine-tune their performance.

10. Research and Development:

- Allocate resources to research emerging technologies and trends in AI and recommendation systems.

- Stay ahead of the curve by experimenting with innovative AI models and techniques.

Conclusion:

Phase 2 aims to take IMDb score prediction to the next level by harnessing cutting-edge technologies and a deep understanding of user behavior and preferences. By focusing on innovation, the platform can provide users with IMDb score predictions that are not only highly accurate but also tailored to their unique tastes and interests.

```
In [6]: import pandas as pd
```

```
In [10]: import matplotlib.pyplot as plt
```

```
In [11]: data=pd.read_csv(r"imdb_top_1000.csv")
```

```
In [12]: print(data)
```

```
                                            Poster_Link  \
0     https://m.media-amazon.com/images/M/MV5BMDFkYT...
1     https://m.media-amazon.com/images/M/MV5BM2MyNj...
2     https://m.media-amazon.com/images/M/MV5BMTMxNT...
3     https://m.media-amazon.com/images/M/MV5BMWMwMG...
4     https://m.media-amazon.com/images/M/MV5BMWU4N2...
..                                                  ...
995   https://m.media-amazon.com/images/M/MV5BNGEwMT...
996   https://m.media-amazon.com/images/M/MV5BODk3Yj...
997   https://m.media-amazon.com/images/M/MV5BM2U3Yz...
998   https://m.media-amazon.com/images/M/MV5BZTBmMj...
999   https://m.media-amazon.com/images/M/MV5BMTY5OD...

                 Series_Title Released_Year Certificate  Runtime  \
0    The Shawshank Redemption          1994           A  142 min
1               The Godfather          1972           A  175 min
2             The Dark Knight          2008          UA  152 min
3      The Godfather: Part II          1974           A  202 min
4                12 Angry Men          1957           U   96 min
..                        ...           ...         ...      ...
995     Breakfast at Tiffany's          1961           A  115 min
996                      Giant          1956           G  201 min
997      From Here to Eternity          1953      Passed  118 min
998                   Lifeboat          1944         NaN   97 min
999               The 39 Steps          1935         NaN   86 min

                        Genre  IMDB_Rating  \
0                       Drama          9.3
1                Crime, Drama          9.2
2        Action, Crime, Drama          9.0
3                Crime, Drama          9.0
4                Crime, Drama          9.0
..                        ...          ...
995   Comedy, Drama, Romance          7.6
996           Drama, Western          7.6
997     Drama, Romance, War          7.6
998              Drama, War          7.6
999  Crime, Mystery, Thriller          7.6
```

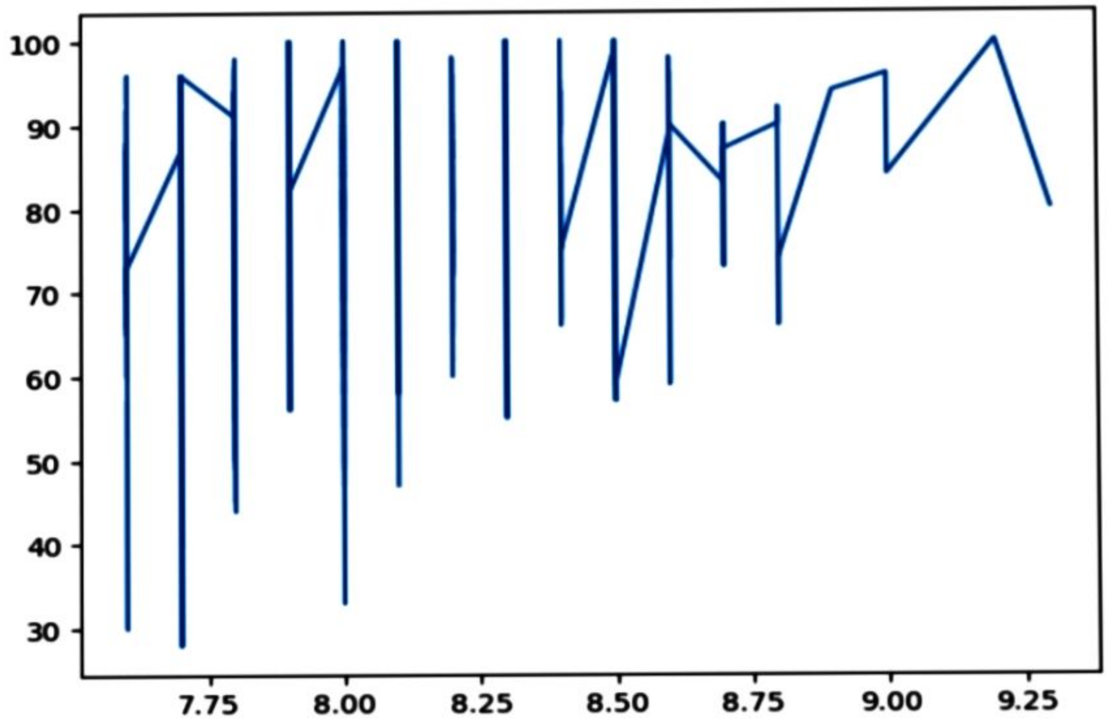
```
                                             Overview  Meta_score  \
0    Two imprisoned men bond over a number of years...        80.0
1    An organized crime dynasty's aging patriarch t...       100.0
2    When the menace known as the Joker wreaks havo...        84.0
3    The early life and career of Vito Corleone in ...        90.0
4    A jury holdout attempts to prevent a miscarria...        96.0
..                                                 ...         ...
995  A young New York socialite becomes interested ...        76.0
996  Sprawling epic covering the life of a Texas ca...        84.0
997  In Hawaii in 1941, a private is cruelly punish...        85.0
998  Several survivors of a torpedoed merchant ship...        78.0
999  A man in London tries to help a counter-espion...        93.0

                 Director             Star1             Star2  \
0          Frank Darabont       Tim Robbins    Morgan Freeman
1    Francis Ford Coppola     Marlon Brando         Al Pacino
2       Christopher Nolan    Christian Bale      Heath Ledger
3    Francis Ford Coppola         Al Pacino   Robert De Niro
4            Sidney Lumet       Henry Fonda       Lee J. Cobb
..                    ...               ...               ...
995         Blake Edwards    Audrey Hepburn    George Peppard
996        George Stevens  Elizabeth Taylor       Rock Hudson
997        Fred Zinnemann    Burt Lancaster  Montgomery Clift
998      Alfred Hitchcock Tallulah Bankhead       John Hodiak
999      Alfred Hitchcock      Robert Donat Madeleine Carroll

               Star3            Star4  No_of_Votes        Gross
0         Bob Gunton   William Sadler      2343110   28,341,469
1         James Caan     Diane Keaton      1620367  134,966,411
2      Aaron Eckhart    Michael Caine      2303232  534,858,444
3      Robert Duvall     Diane Keaton      1129952   57,300,000
4     Martin Balsam      John Fiedler       689845    4,360,000
..               ...              ...          ...          ...
995    Patricia Neal      Buddy Ebsen       166544          NaN
996       James Dean     Carroll Baker        34075          NaN
997     Deborah Kerr       Donna Reed        43374   30,500,000
998    Walter Slezak   William Bendix        26471          NaN
999   Lucie Mannheim   Godfrey Tearle        51853          NaN

[1000 rows x 16 columns]
```

```python
x=(data['IMDB_Rating'])
```

```python
y=(data['Meta_score'])
```

```python
plt.plot(x,y)
```

```
[<matplotlib.lines.Line2D at 0x1e300df36d0>]
```

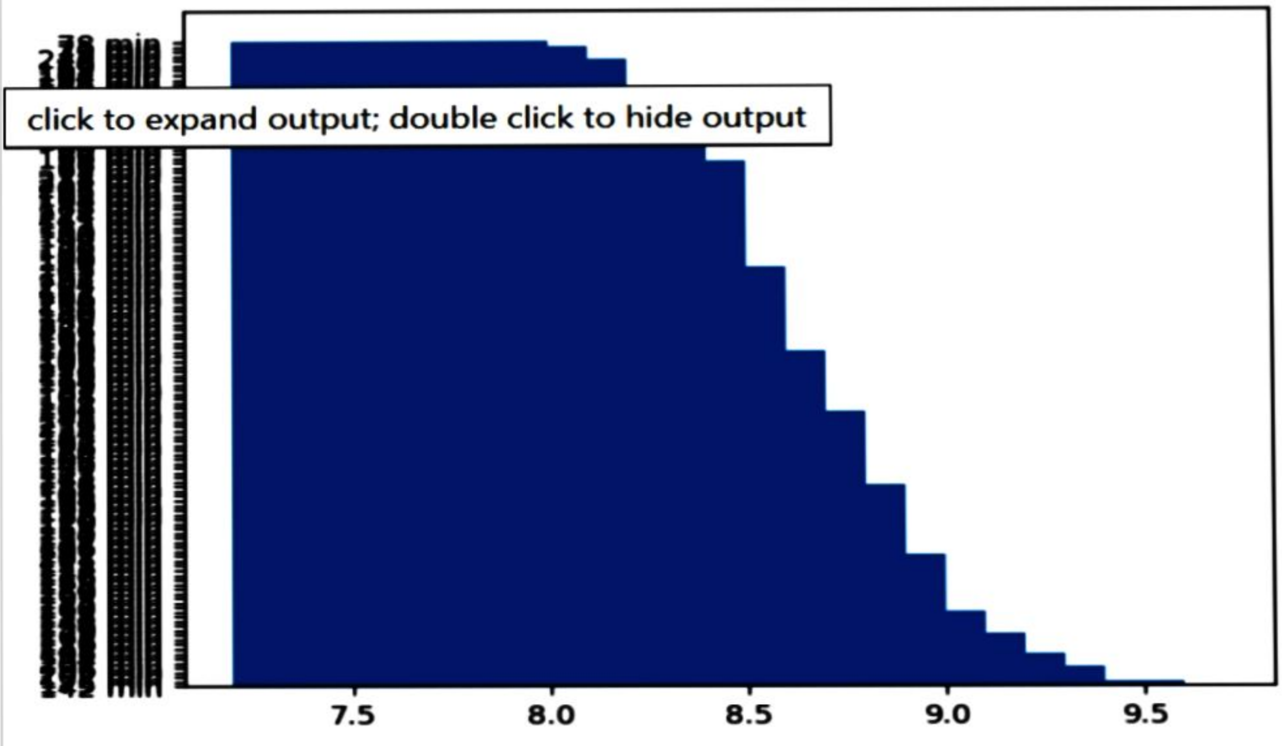



```python
x=(data['IMDB_Rating'])
```

```python
y=(data['Meta_score'])
```

```python
plt.bar(x,y)
```

```
<BarContainer object of 1000 artists>
```

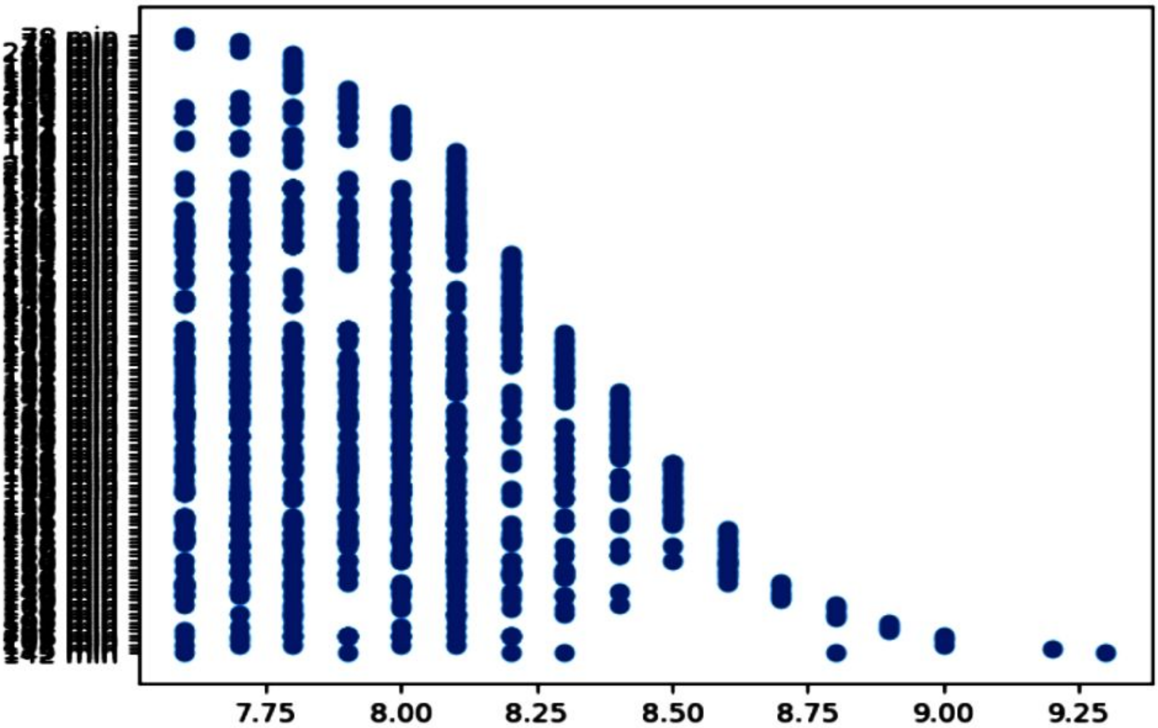


```
]: x=(data['IMDB_Rating'])
```

```
]: y=(data['Meta_score'])
```

```
]: plt.scatter(x,y)
```

```
]: <matplotlib.collections.PathCollection at 0x1e303ddefa0>
```

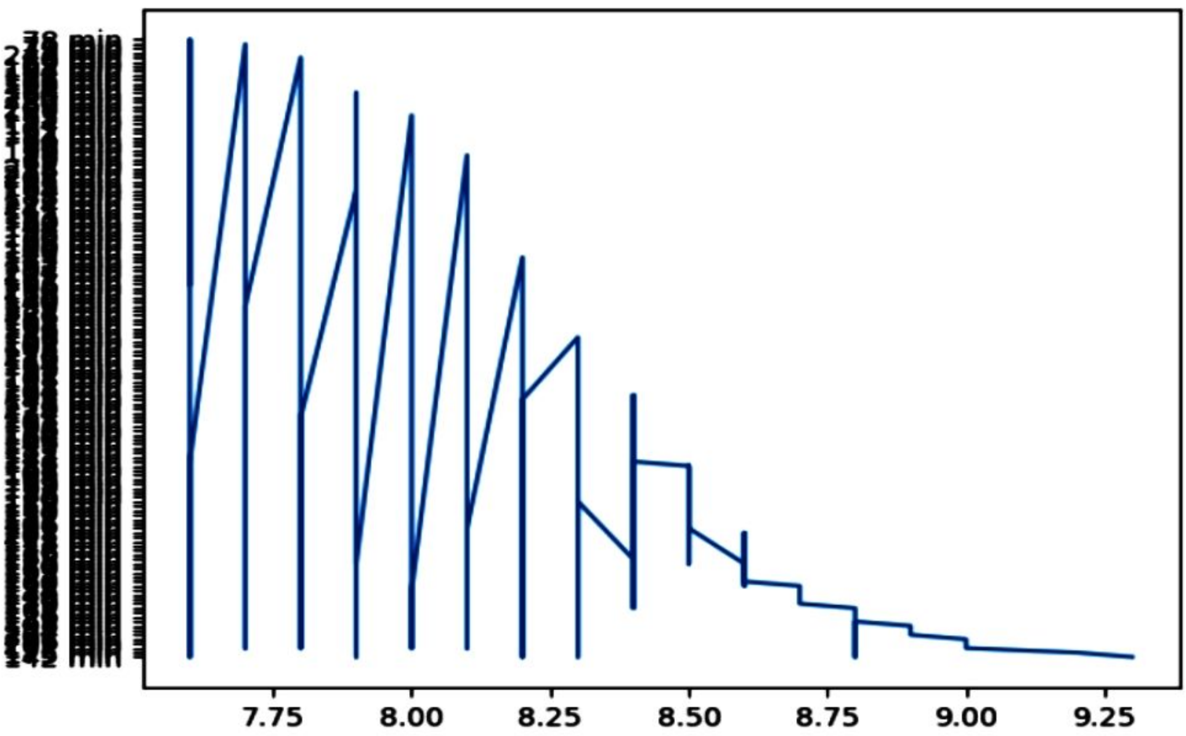



```
x=(data['Released_Year'])
```

```
y=(data['Runtime'])
```

```
plt.plot(x,y)
```

```
[<matplotlib.lines.Line2D at 0x1e303f5b640>]
```

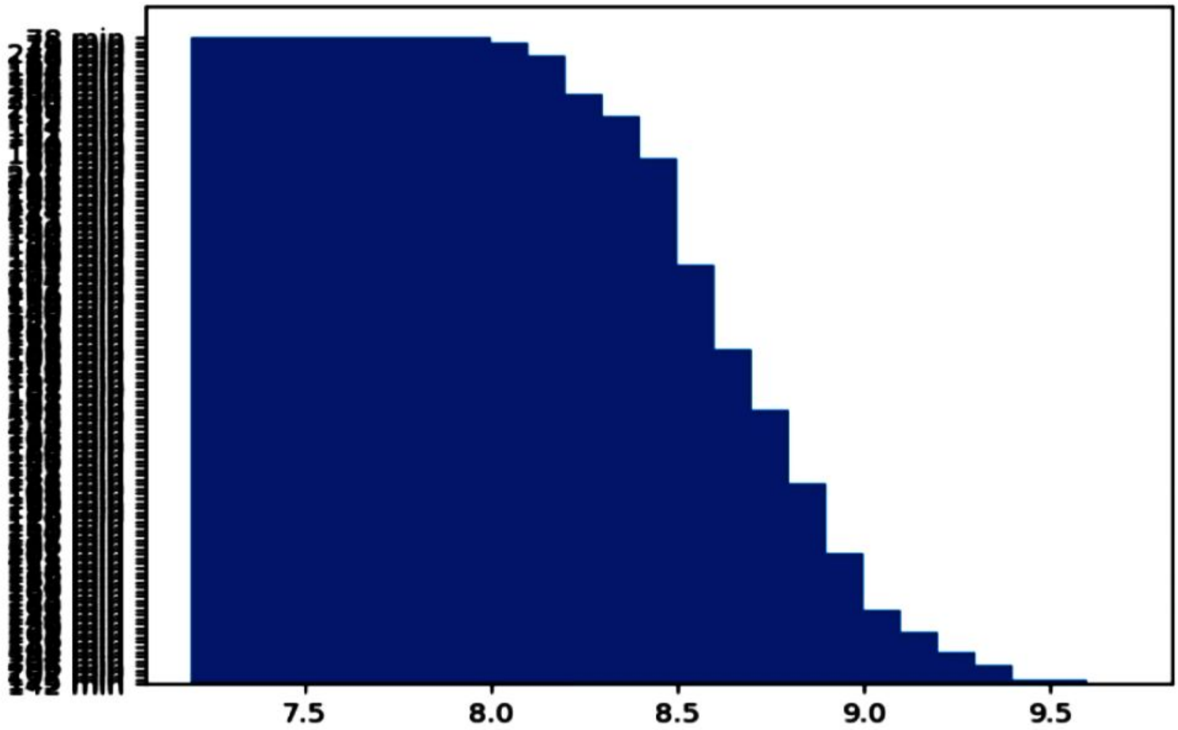


]: `x=(data['Released_Year'])`

]: `y=(data['Runtime'])`

]: `plt.bar(x,y)`

]: <BarContainer object of 1000 artists>

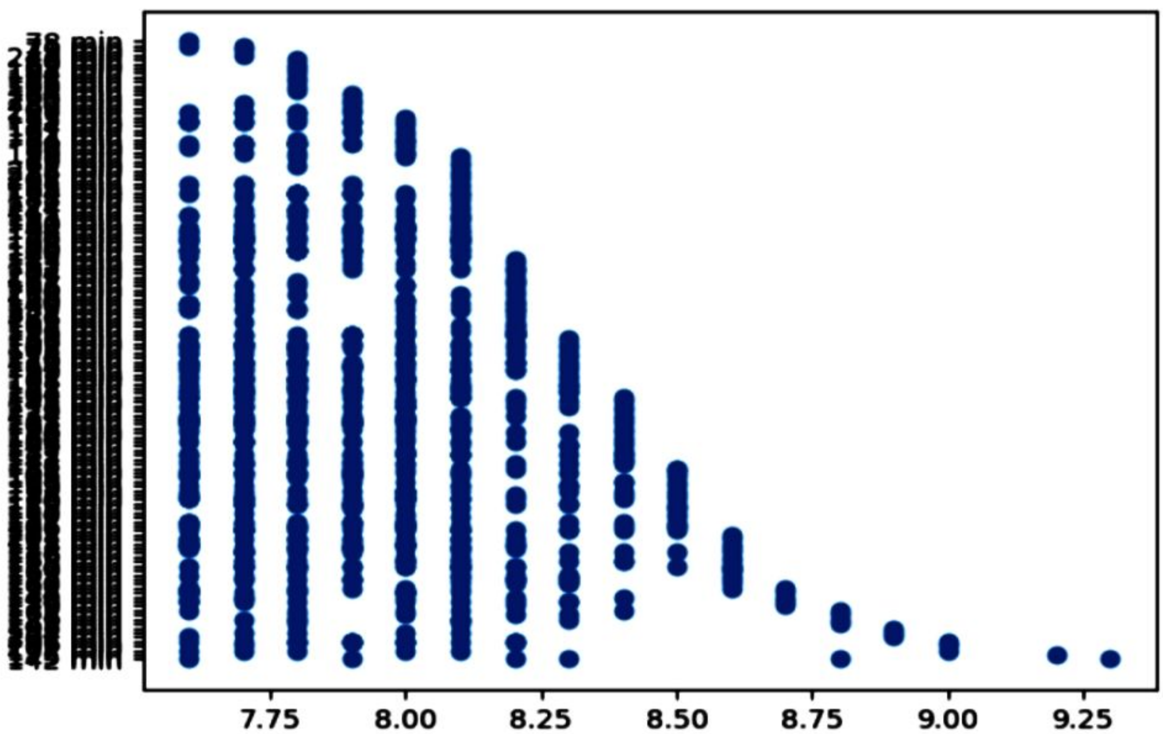


: `x=(data['Released_Year'])`

: `y=(data['Runtime'])`

: `plt.scatter(x,y)`

: <matplotlib.collections.PathCollection at 0x1e305b1aac0>



```python
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
```

```python
data = pd.read_csv(r"imdb_top_1000 - Copy.csv")
print(data.head())
```

```
                                         Poster_Link  \
0  https://m.media-amazon.com/images/M/MV5BMDFkYT...
1  https://m.media-amazon.com/images/M/MV5BM2MyNj...
2  https://m.media-amazon.com/images/M/MV5BMTMxNT...
3  https://m.media-amazon.com/images/M/MV5BMWMwMG...
4  https://m.media-amazon.com/images/M/MV5BMWU4N2...

              Series_Title Released_Year Runtime                 Genre  \
0  The Shawshank Redemption          1994 142 min                 Drama
1             The Godfather          1972 175 min          Crime, Drama
2           The Dark Knight          2008 152 min  Action, Crime, Drama
3    The Godfather: Part II          1974 202 min          Crime, Drama
4              12 Angry Men          1957  96 min          Crime, Drama

   IMDB_Rating                                           Overview  Meta_score  \
0          9.3  Two imprisoned men bond over a number of years...          80
1          9.2  An organized crime dynasty's aging patriarch t...         100
2          9.0  When the menace known as the Joker wreaks havo...          84
3          9.0  The early life and career of Vito Corleone in ...          90
4          9.0  A jury holdout attempts to prevent a miscarria...          96

               Director           Star1            Star2          Star3  \
0        Frank Darabont     Tim Robbins   Morgan Freeman     Bob Gunton
1  Francis Ford Coppola   Marlon Brando       Al Pacino      James Caan
2     Christopher Nolan  Christian Bale     Heath Ledger  Aaron Eckhart
3  Francis Ford Coppola       Al Pacino  Robert De Niro  Robert Duvall
4          Sidney Lumet     Henry Fonda      Lee J. Cobb  Martin Balsam

           Star4  No_of_Votes        Gross Certificate  Label
0 William Sadler      2343110  2,83,41,469           A      1
1   Diane Keaton      1620367 13,49,66,411           A      1
2  Michael Caine      2303232 53,48,58,444          UA      2
3   Diane Keaton      1129952  5,73,00,000           A      1
4   John Fiedler       689845    43,60,000           U      3
```

```python
print(data.columns)

feature=data[['Meta_score','IMDB_Rating' ]]

#independent var
x=np.asarray(feature)

#dependent var
y=np.asarray(data['Label'])
```

```
Index(['Poster_Link', 'Series_Title', 'Released_Year', 'Runtime', 'Genre',
       'IMDB_Rating', 'Overview', 'Meta_score', 'Director', 'Star1', 'Star2',
       'Star3', 'Star4', 'No_of_Votes', 'Gross', 'Certificate', 'Label'],
      dtype='object')
```

```python
from sklearn.model_selection import train_test_split
from sklearn import tree
clf = tree.DecisionTreeClassifier()

clf.fit(x_train, y_train)
#decision tree

y_predict2=clf.predict(x_test)
x_train, x_test , y_train, y_test = train_test_split(x, y, test_size=0.3, random_state=1 )
```

```python
from sklearn.metrics import confusion_matrix

cm = confusion_matrix(y_test, y_predict2)
print(cm)
```

```
[[4 0 5]
 [3 1 0]
 [4 0 1]]
```

```python
from sklearn.metrics import accuracy_score

accuracy = accuracy_score(y_test, y_predict)
print('Accuracy (Linear Kernel): ', "%.2f" % (accuracy*100))
```

```
Accuracy (Linear Kernel):  50.00
```

```python
from sklearn.metrics import precision_score
from sklearn.metrics import recall_score
#calculating precision and reall
precision = precision_score(y_test, y_predict2, average='weighted')
recall = recall_score(y_test, y_predict2, average='weighted')
print('Precision: ', "%.2f" % (precision*100))
print('Recall: ', "%.2f" % (recall*100))
```

```
Precision:  45.03
Recall:  33.33
```