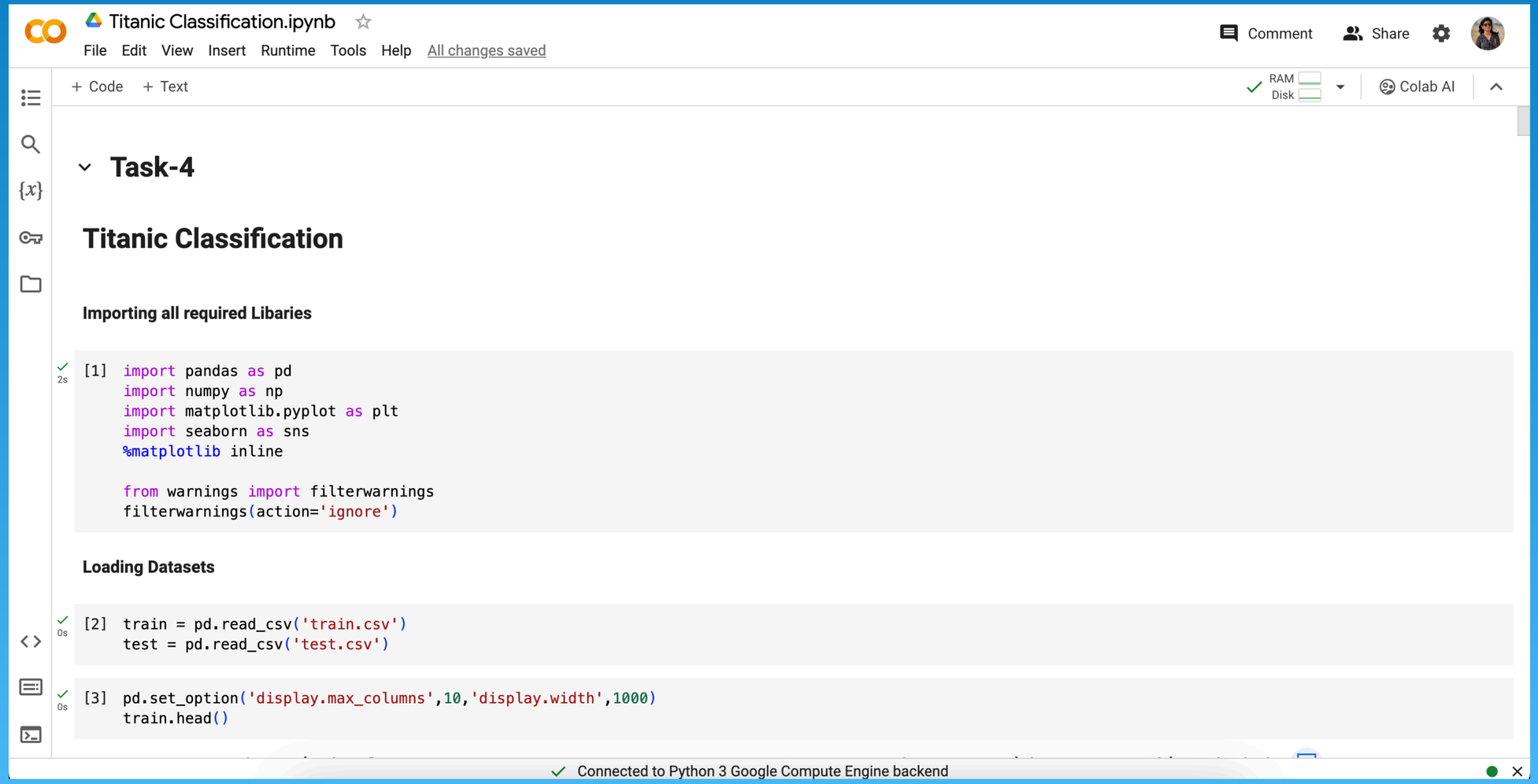
Task-4

Titanic Classification

Samritha A.R

Importing all necessary Libraries & Loading Datasets



Fetching 1st 5 rows of train & test datasets

Titanic Classification.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

RAM Disk Colab AI

```
[3] pd.set_option('display.max_columns',10,'display.width',1000)
    train.head()
```

	PassengerId	Survived	Pclass	Name	Sex	...	Parch	Ticket	Fare	Cabin	Embarked
0	1	0	3	Braund, Mr. Owen Harris	male	...	0	A/5 21171	7.2500	NaN	S
1	2	1	1	Cumings, Mrs. John Bradley (Florence Briggs Th...	female	...	0	PC 17599	71.2833	C85	C
2	3	1	3	Heikkinen, Miss. Laina	female	...	0	STON/O2. 3101282	7.9250	NaN	S
3	4	1	1	Futrelle, Mrs. Jacques Heath (Lily May Peel)	female	...	0	113803	53.1000	C123	S
4	5	0	3	Allen, Mr. William Henry	male	...	0	373450	8.0500	NaN	S

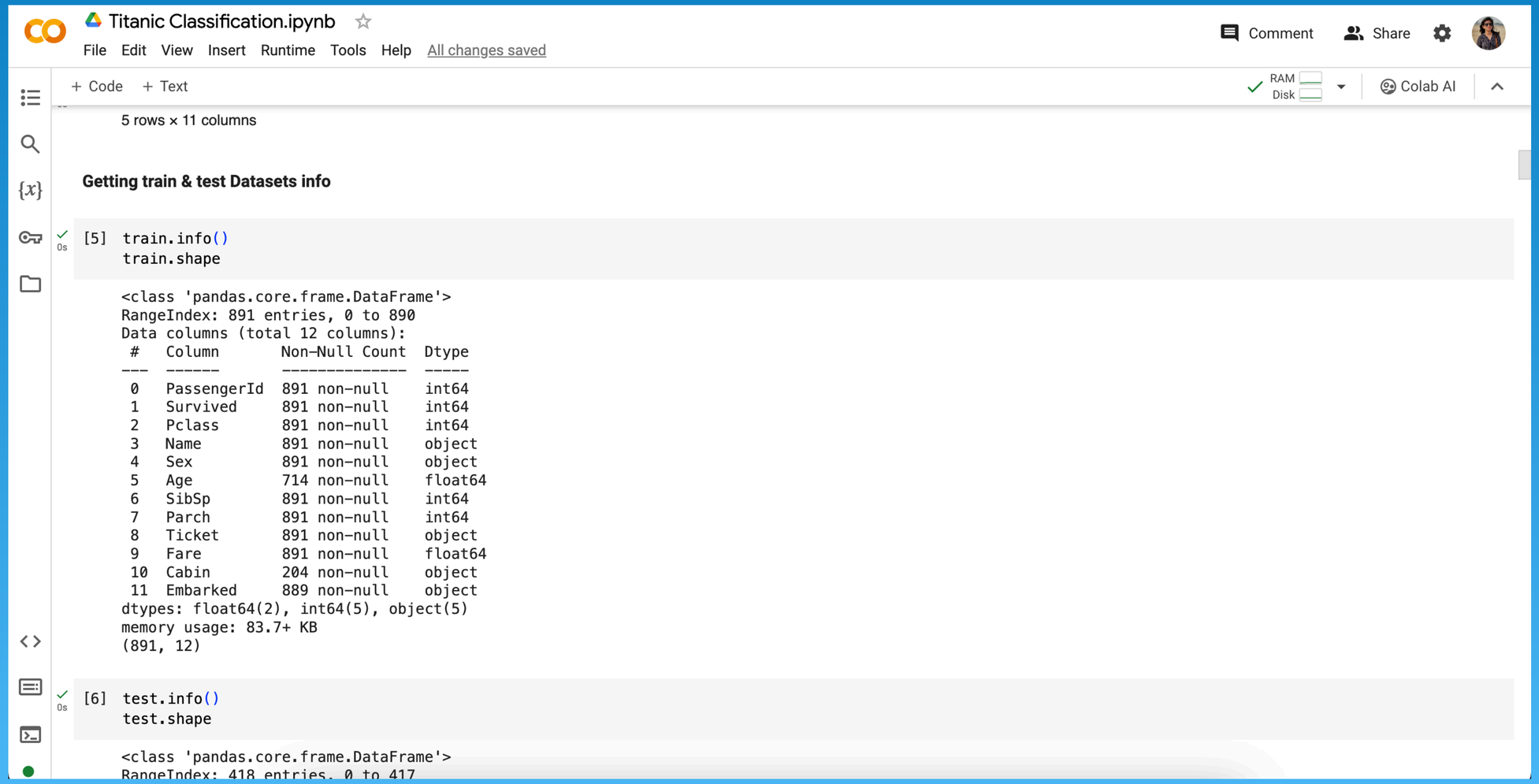
5 rows × 12 columns

```
[4] pd.set_option('display.max_columns',10,'display.width',1000)
    test.head()
```

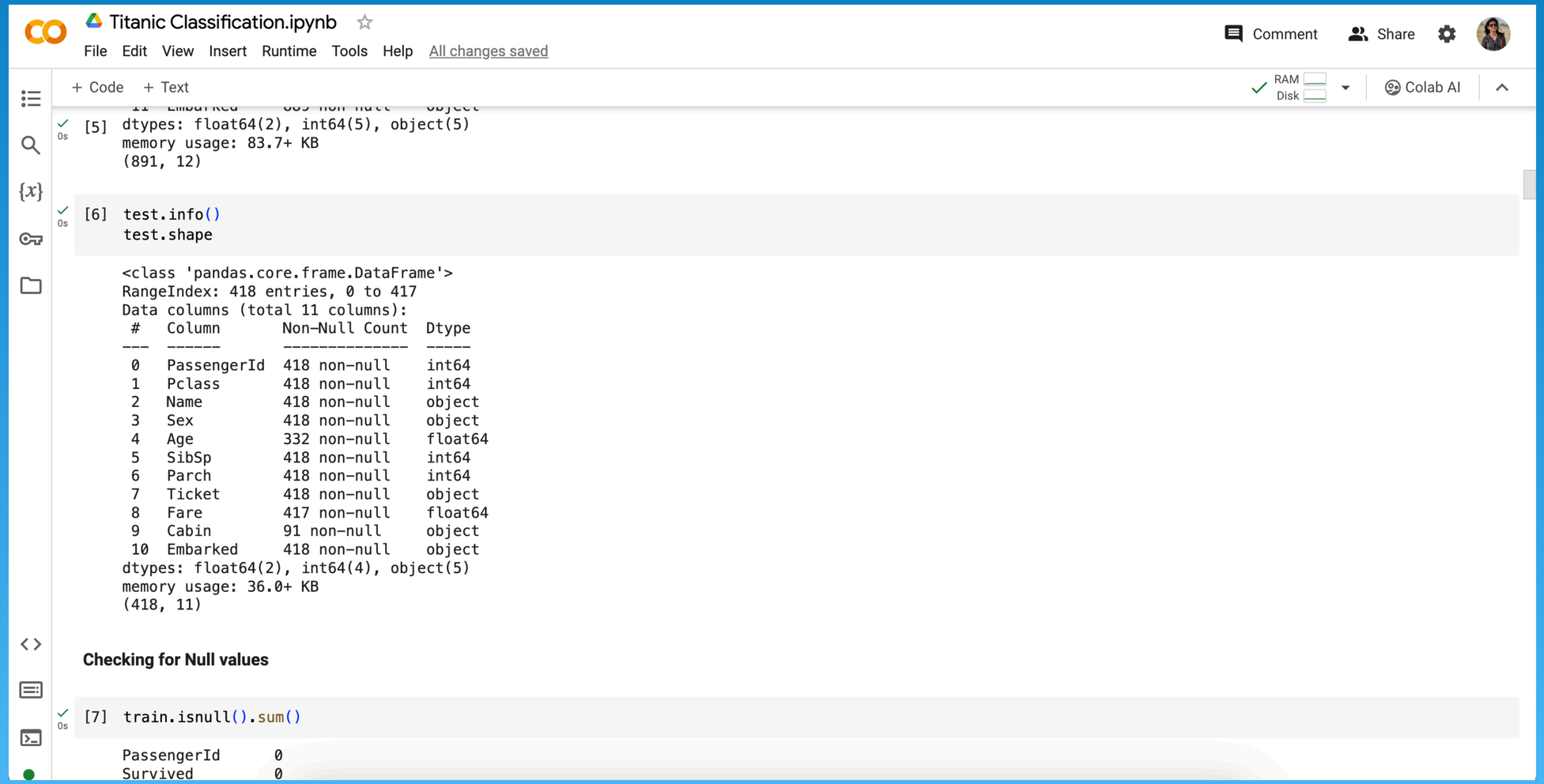
	PassengerId	Pclass	Name	Sex	Age	...	Parch	Ticket	Fare	Cabin	Embarked
0	892	3	Kelly, Mr. James	male	34.5	...	0	330911	7.8292	NaN	Q
1	893	3	Wilkes, Mrs. James (Ellen Needs)	female	47.0	...	0	363272	7.0000	NaN	S
2	894	2	Myles, Mr. Thomas Francis	male	62.0	...	0	240276	9.6875	NaN	Q
3	895	3	Wirz, Mr. Albert	male	27.0	...	0	315154	8.6625	NaN	S
4	896	3	Hirvonen, Mrs. Alexander (Helga E Lindqvist)	female	22.0	...	1	3101298	12.2875	NaN	S

5 rows × 11 columns

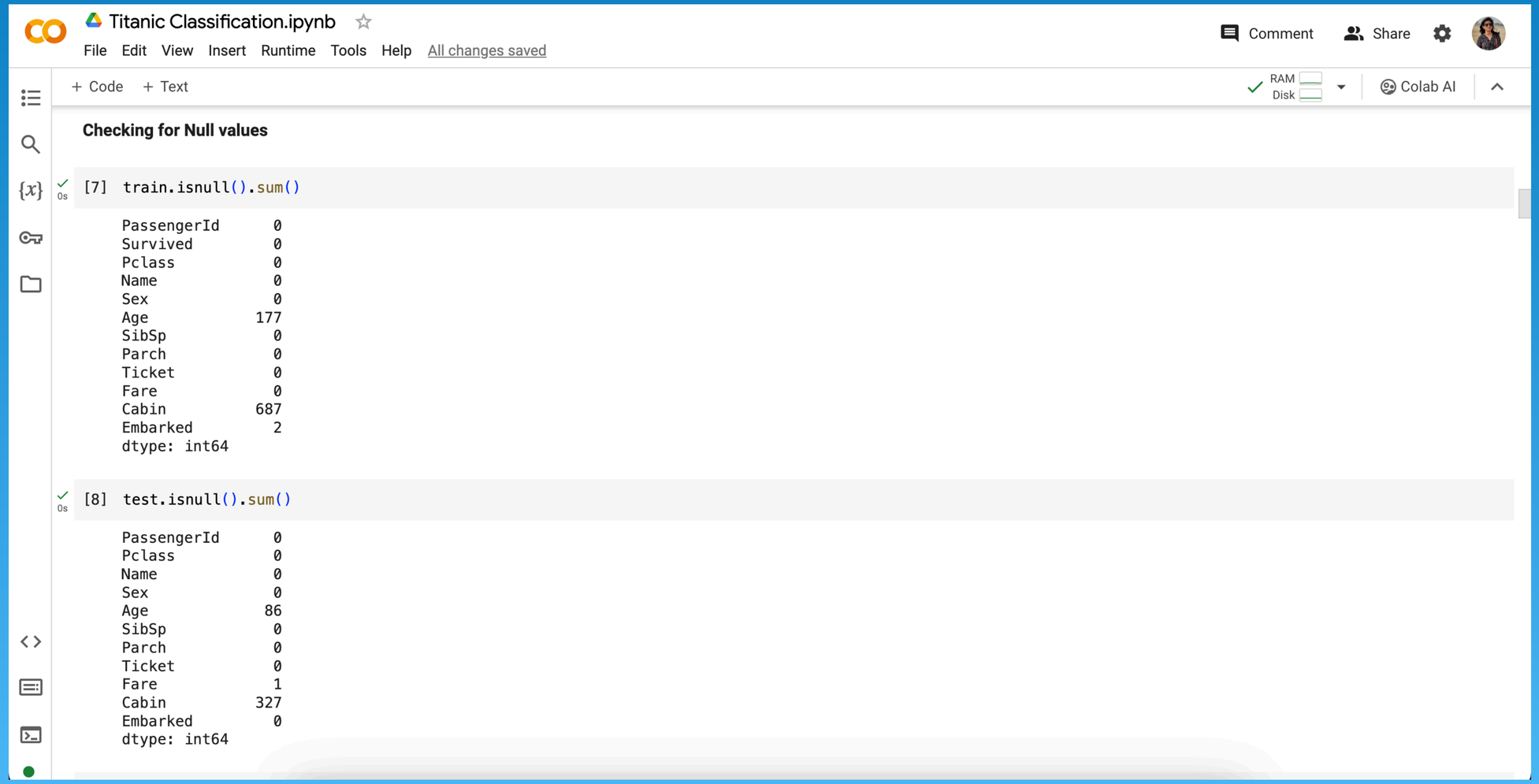
Getting Train dataset Info & Shape



Getting Test dataset Info & Shape



Checking for Null values



Heatmap for Train set to check null values

Titanic Classification.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

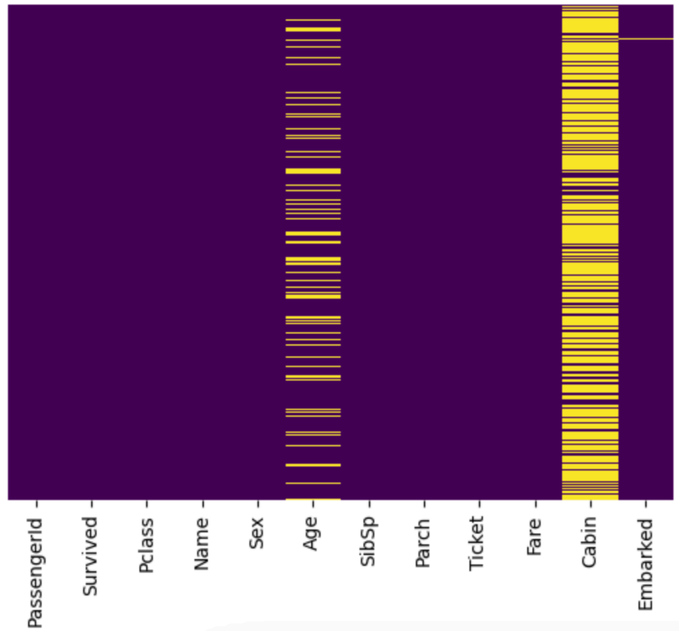
+ Code + Text

RAM Disk Colab AI

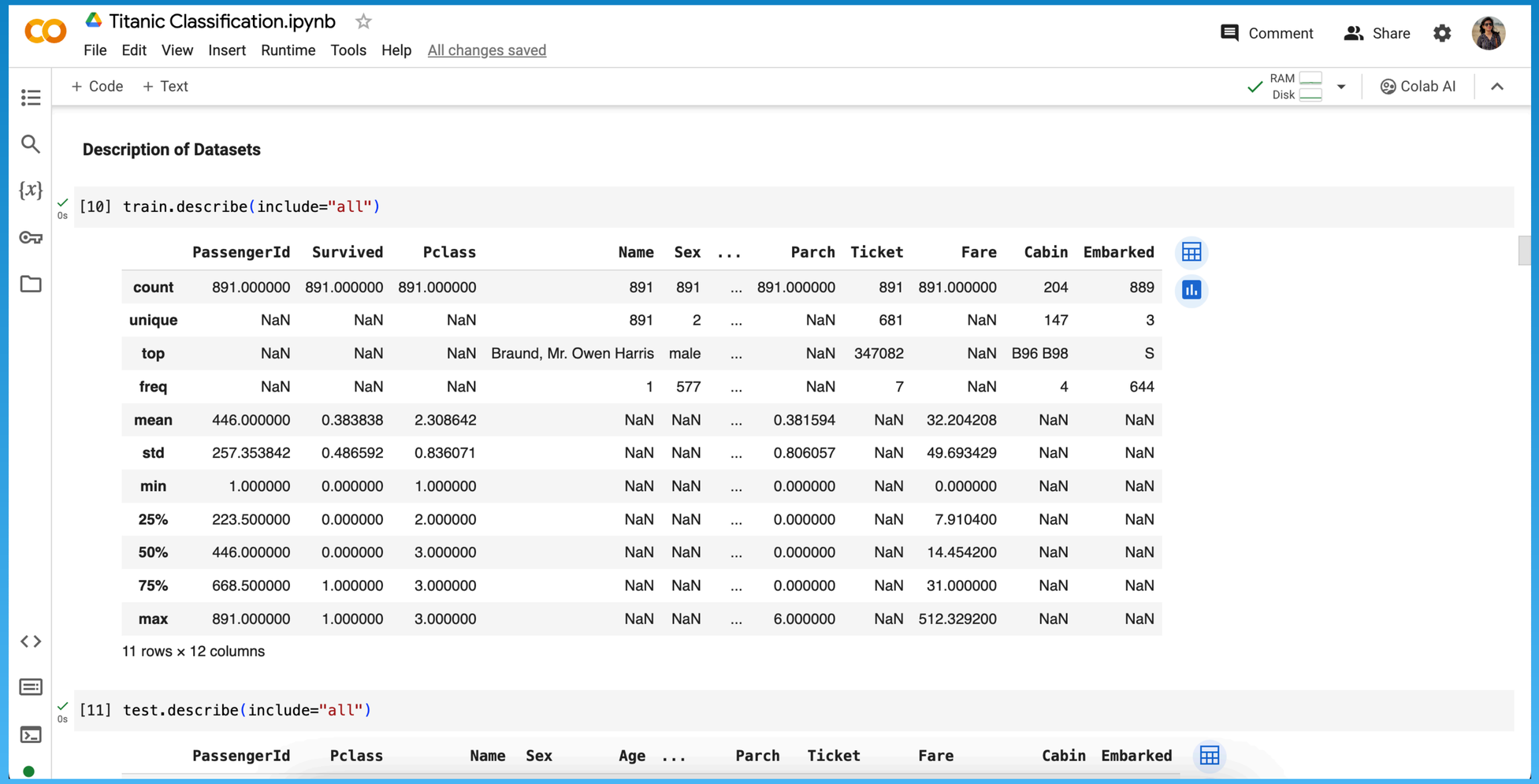
```
Embarked         0
dtype: int64
```

```
[9]  sns.heatmap(train.isnull(),yticklabels=False,cbar=False,cmap='viridis')
```

```
<Axes: >
```



Description of Train set

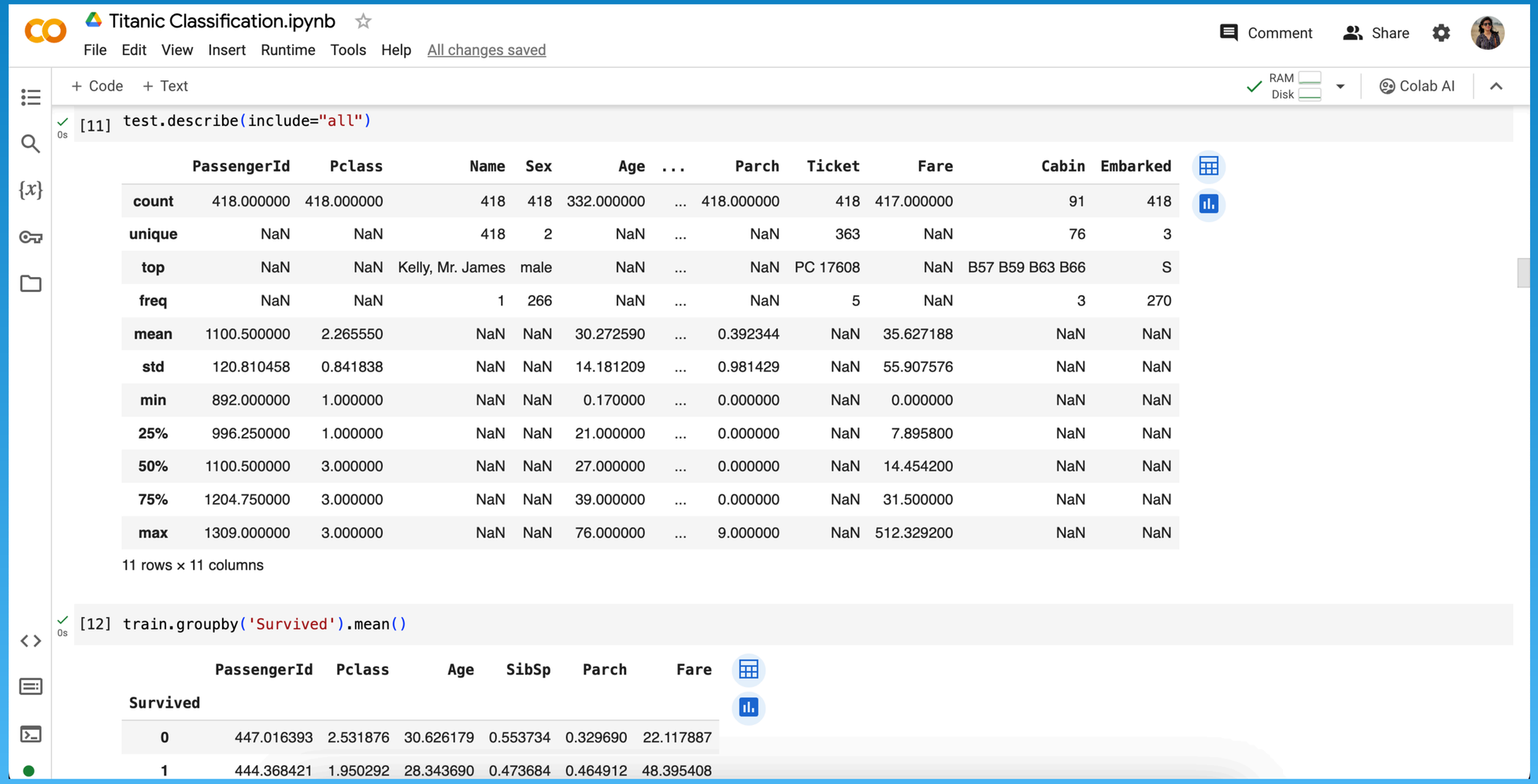


	PassengerId	Survived	Pclass	Name	Sex	...	Parch	Ticket	Fare	Cabin	Embarked
count	891.000000	891.000000	891.000000	891	891	...	891.000000	891	891.000000	204	889
unique	NaN	NaN	NaN	891	2	...	NaN	681	NaN	147	3
top	NaN	NaN	NaN	Braund, Mr. Owen Harris	male	...	NaN	347082	NaN	B96 B98	S
freq	NaN	NaN	NaN	1	577	...	NaN	7	NaN	4	644
mean	446.000000	0.383838	2.308642	NaN	NaN	...	0.381594	NaN	32.204208	NaN	NaN
std	257.353842	0.486592	0.836071	NaN	NaN	...	0.806057	NaN	49.693429	NaN	NaN
min	1.000000	0.000000	1.000000	NaN	NaN	...	0.000000	NaN	0.000000	NaN	NaN
25%	223.500000	0.000000	2.000000	NaN	NaN	...	0.000000	NaN	7.910400	NaN	NaN
50%	446.000000	0.000000	3.000000	NaN	NaN	...	0.000000	NaN	14.454200	NaN	NaN
75%	668.500000	1.000000	3.000000	NaN	NaN	...	0.000000	NaN	31.000000	NaN	NaN
max	891.000000	1.000000	3.000000	NaN	NaN	...	6.000000	NaN	512.329200	NaN	NaN

11 rows × 12 columns

	PassengerId	Pclass	Name	Sex	Age	...	Parch	Ticket	Fare	Cabin	Embarked

Description of Train set



	PassengerId	Pclass	Name	Sex	Age	...	Parch	Ticket	Fare	Cabin	Embarked
count	418.000000	418.000000	418	418	332.000000	...	418.000000	418	417.000000	91	418
unique	NaN	NaN	418	2	NaN	...	NaN	363	NaN	76	3
top	NaN	NaN	Kelly, Mr. James	male	NaN	...	NaN	PC 17608	NaN	B57 B59 B63 B66	S
freq	NaN	NaN	1	266	NaN	...	NaN	5	NaN	3	270
mean	1100.500000	2.265550	NaN	NaN	30.272590	...	0.392344	NaN	35.627188	NaN	NaN
std	120.810458	0.841838	NaN	NaN	14.181209	...	0.981429	NaN	55.907576	NaN	NaN
min	892.000000	1.000000	NaN	NaN	0.170000	...	0.000000	NaN	0.000000	NaN	NaN
25%	996.250000	1.000000	NaN	NaN	21.000000	...	0.000000	NaN	7.895800	NaN	NaN
50%	1100.500000	3.000000	NaN	NaN	27.000000	...	0.000000	NaN	14.454200	NaN	NaN
75%	1204.750000	3.000000	NaN	NaN	39.000000	...	0.000000	NaN	31.500000	NaN	NaN
max	1309.000000	3.000000	NaN	NaN	76.000000	...	9.000000	NaN	512.329200	NaN	NaN

11 rows × 11 columns

	PassengerId	Pclass	Age	SibSp	Parch	Fare
Survived						
0	447.016393	2.531876	30.626179	0.553734	0.329690	22.117887
1	444.368421	1.950292	28.343690	0.473684	0.464912	48.395408

Correlation Plot of Train dataset

Titanic Classification.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

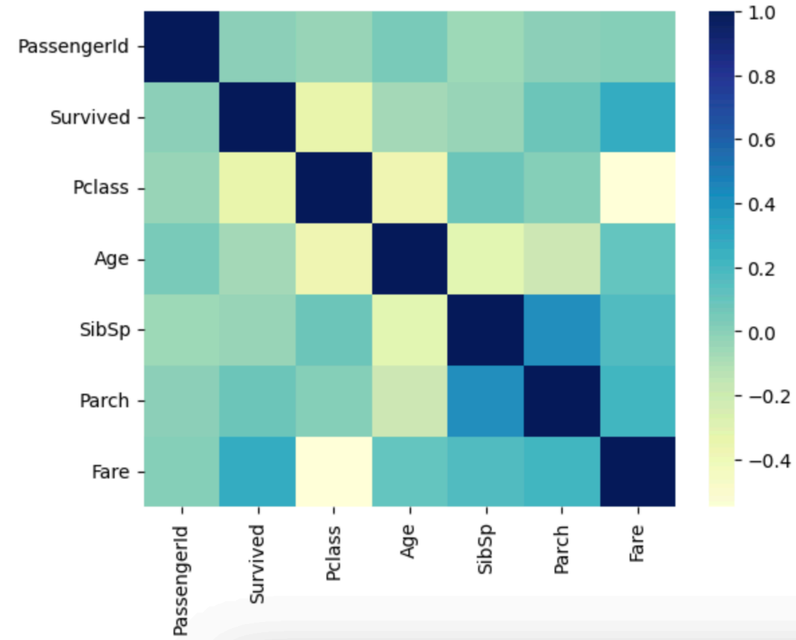
+ Code + Text

RAM Disk Colab AI

Correlation Plot of train dataset

```
[13] sns.heatmap(train.corr(), cmap="YlGnBu")
     plt.show
```

```
<function matplotlib.pyplot.show(close=None, block=None)>
```



Exploratory Data Analysis for Train data

Titanic Classification.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

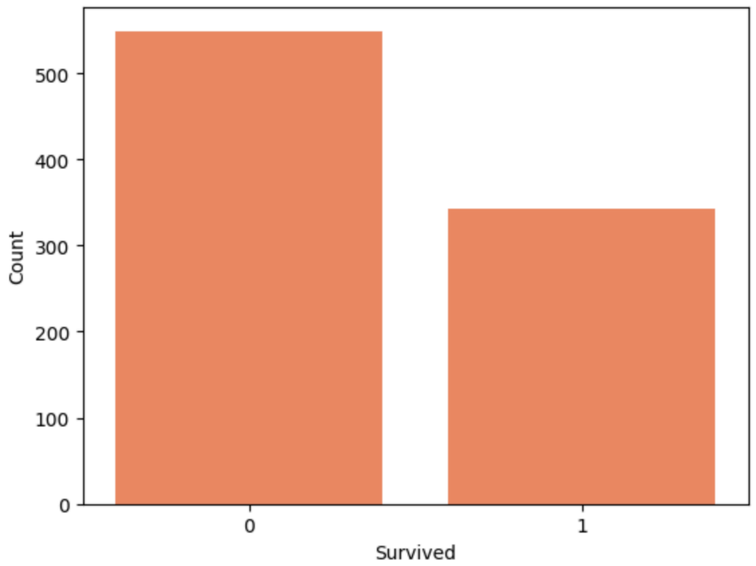
+ Code + Text

RAM Disk Colab AI

Exploratory Data Analysis

```
[14] sns.countplot(x='Survived',data=train, color="coral")
     plt.ylabel('Count')
     plt.show()
```

0s




```
[15] sns.countplot(x='Survived', hue='Sex', data=train, palette='jet_r')
```

Exploratory Data Analysis for Train data

Titanic Classification.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

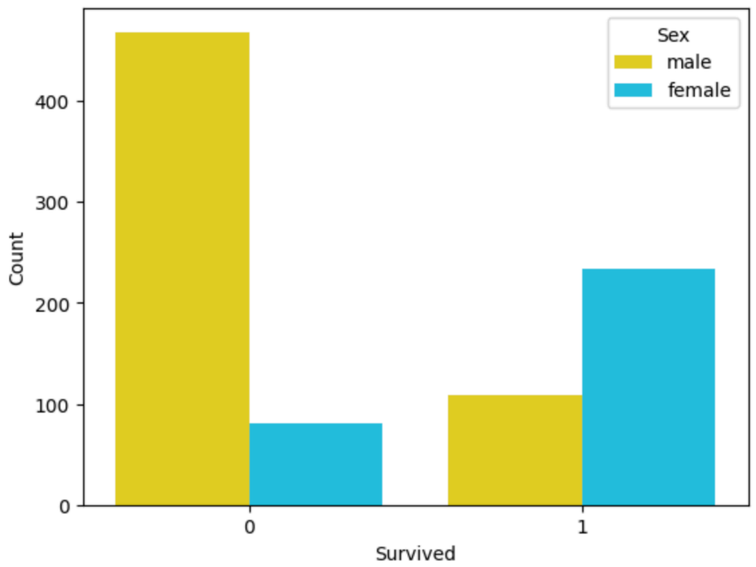
RAM Disk Colab AI

[14]

0 1

Survived

```
sns.countplot(x='Survived',hue='Sex',data=train,palette='jet_r')
plt.ylabel('Count')
plt.show()
```



[16] sns.set_style('whitegrid')

Exploratory Data Analysis for Train data

Titanic Classification.ipynb

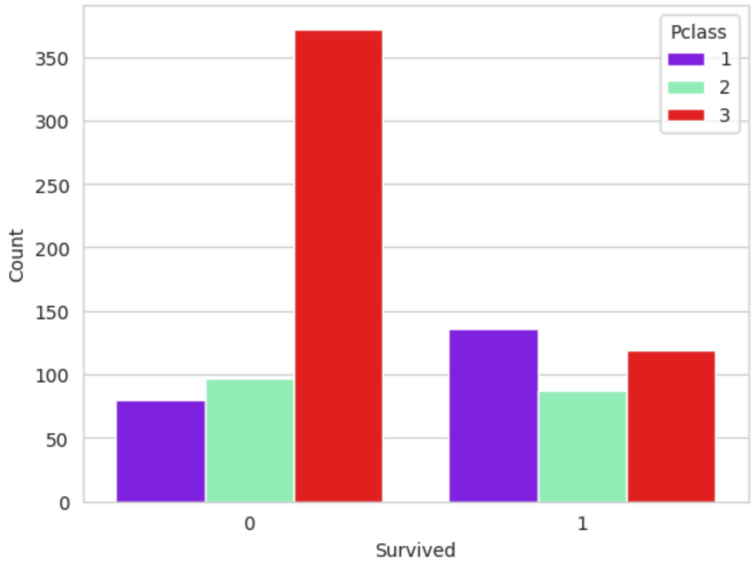
File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

RAM Disk Colab AI

```
[16] sns.set_style('whitegrid')
     sns.countplot(x='Survived',hue='Pclass',data=train,palette='rainbow')
     plt.ylabel('Count')
     plt.show()
```




```
[17] sns.distplot(train['Age'].dropna(),kde=False,color='darkred',bins=40)
```

Exploratory Data Analysis for Train data

Titanic Classification.ipynb

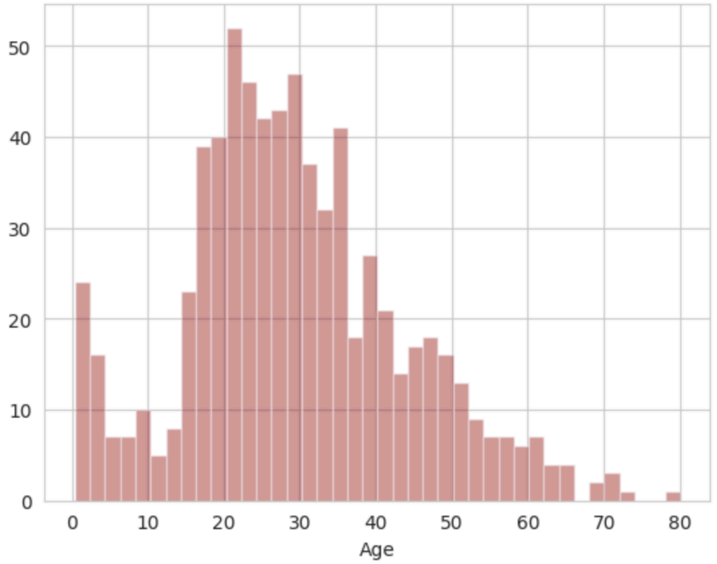
File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

RAM Disk Colab AI

```
[17] sns.distplot(train['Age'].dropna(),kde=False,color='darkred',bins=40)
     plt.show()
```




```
[18] sns.countplot(x='SibSp',data=train)
     plt.ylabel('Count')
     plt.show()
```

Exploratory Data Analysis for Train data

Titanic Classification.ipynb

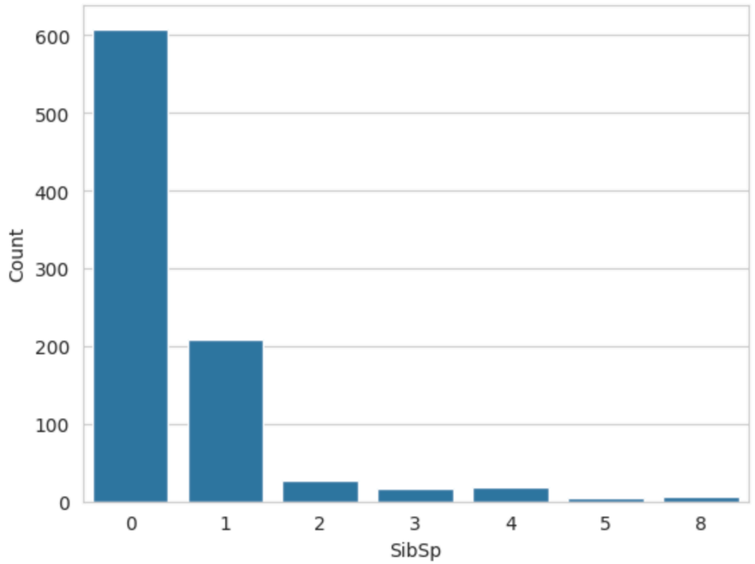
File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

RAM Disk Colab AI

```
[18] sns.countplot(x='SibSp',data=train)
     plt.ylabel('Count')
     plt.show()
```




```
[19] train['Fare'].hist(color='green',bins=40,figsize=(8,4))
     plt.show()
```

Exploratory Data Analysis for Train data

Titanic Classification.ipynb

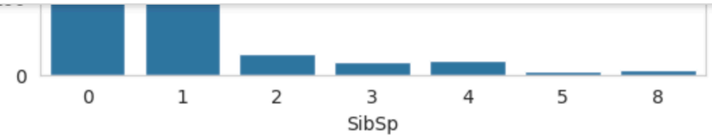
File Edit View Insert Runtime Tools Help All changes saved

Comment Share

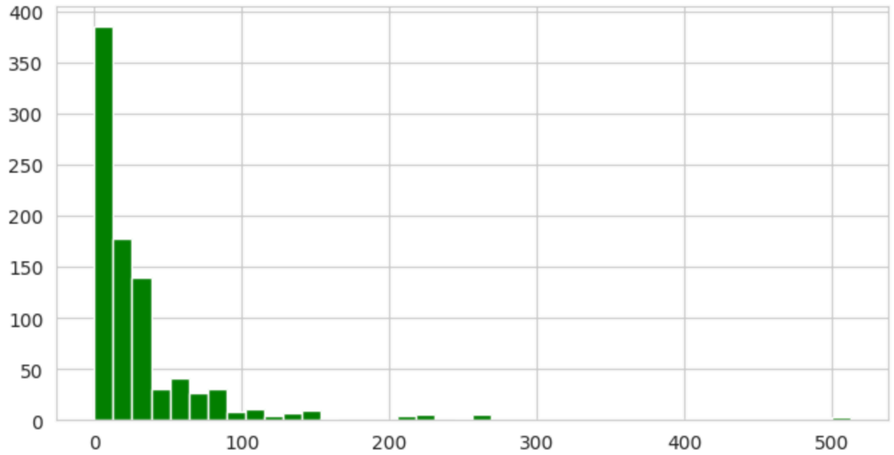
+ Code + Text

RAM Disk Colab AI

[18]



[19]
```
train['Fare'].hist(color='green',bins=40,figsize=(8,4))
plt.show()
```



[20]
```
fig = plt.figure()
ax = fig.add_axes([0,0,1,1])
ax.axis('equal')
```

Exploratory Data Analysis for Train data

Titanic Classification.ipynb

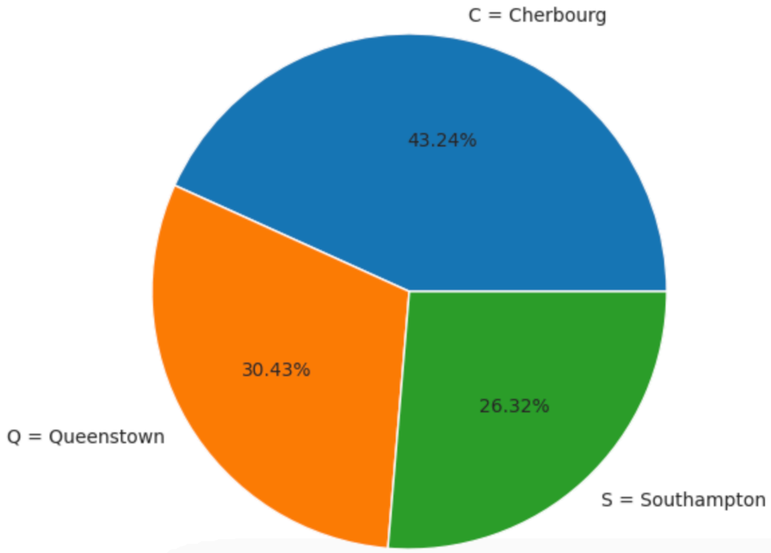
File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

RAM Disk Colab AI

```
[20] fig = plt.figure()
     ax = fig.add_axes([0,0,1,1])
     ax.axis('equal')
     l = ['C = Cherbourg', 'Q = Queenstown', 'S = Southampton']
     s = [0.553571,0.389610,0.336957]
     ax.pie(s, labels = l,autopct='%1.2f%%')
     plt.show()
```



Exploratory Data Analysis for Train data

Titanic Classification.ipynb

File Edit View Insert Runtime Tools Help All changes saved

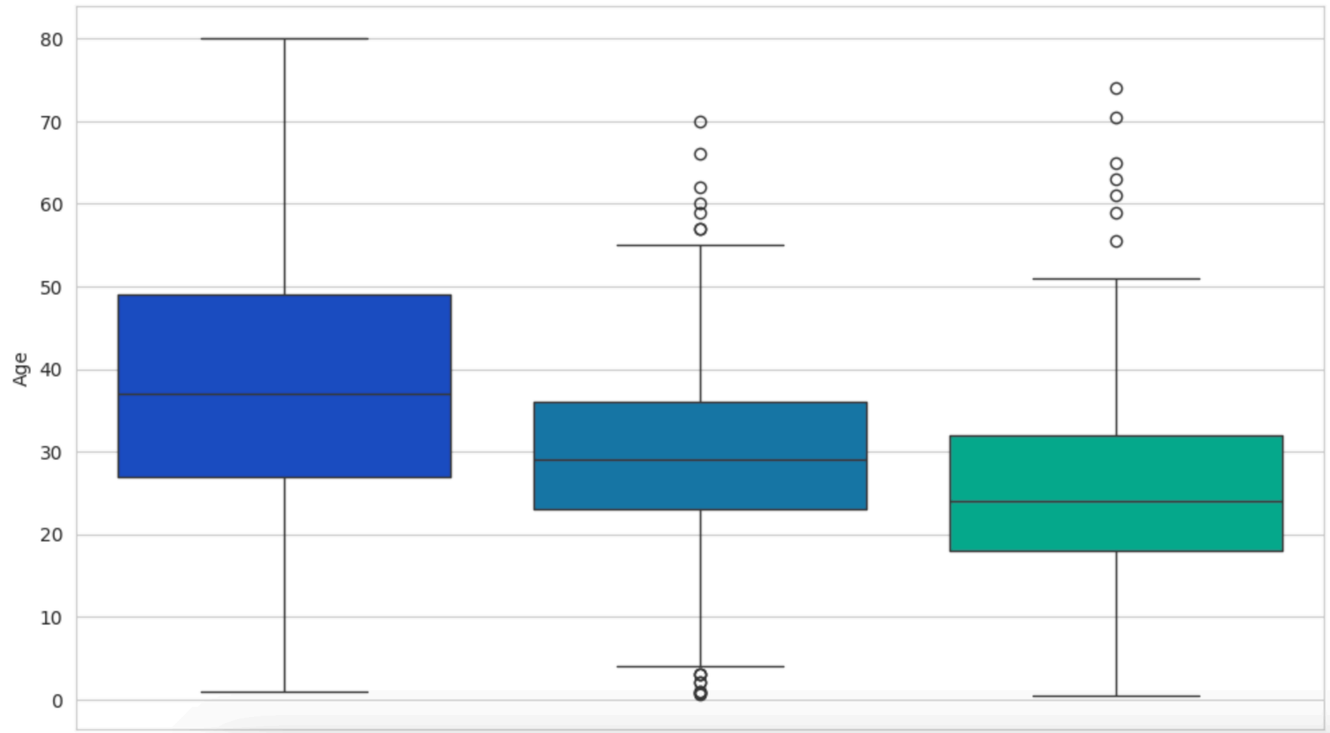
Comment Share

+ Code + Text

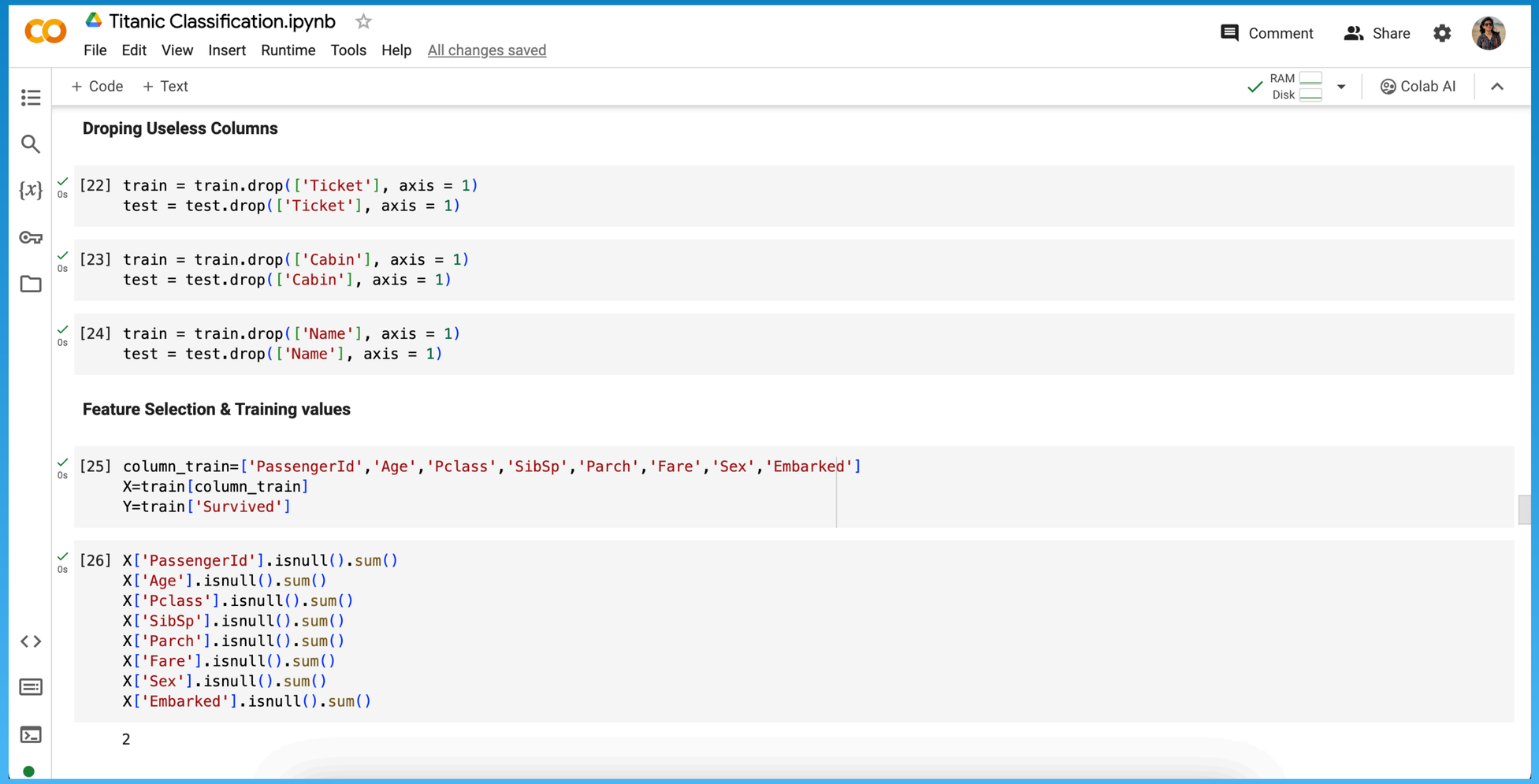
RAM Disk Colab AI

```
[21] plt.figure(figsize=(12, 7))
     sns.boxplot(x='Pclass',y='Age',data=train,palette='winter')
```

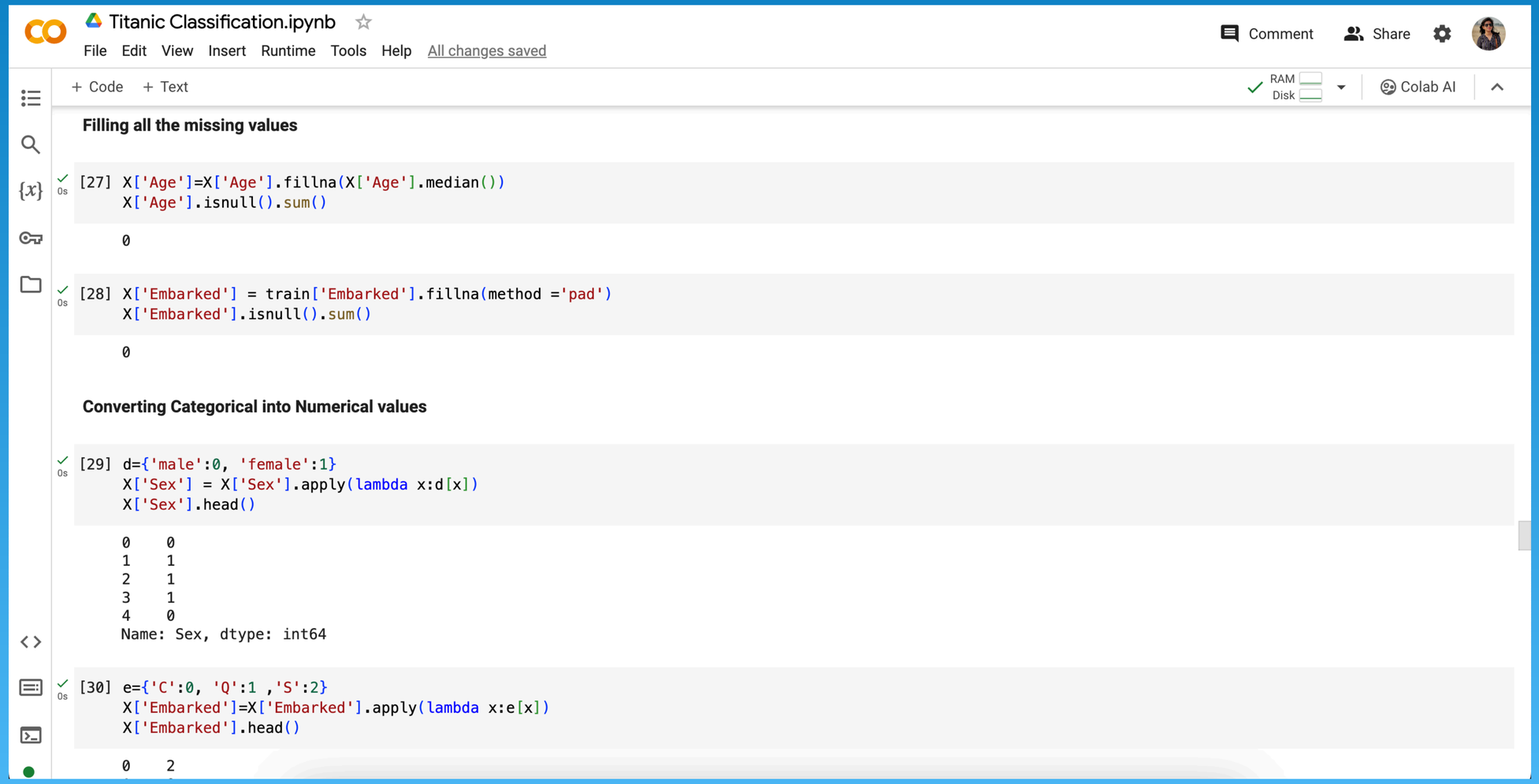
```
<Axes: xlabel='Pclass', ylabel='Age'>
```



Dropping Useless Columns & Feature Selecting Training values



Filling missing values & Converting Categorical to Numerical values



Cross-checking for null values and its datatypes

Titanic Classification.ipynb

File Edit View Insert Runtime Tools Help All changes saved

Comment Share

+ Code + Text

RAM Disk Colab AI

```
[30] e={'C':0, 'Q':1 ,'S':2}
     X['Embarked']=X['Embarked'].apply(lambda x:e[x])
     X['Embarked'].head()
```

```
0    2
1    0
2    2
3    2
4    2
Name: Embarked, dtype: int64
```

```
[31] X.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 891 entries, 0 to 890
Data columns (total 8 columns):
 #   Column       Non-Null Count  Dtype
---  ------       --------------  -----
 0   PassengerId  891 non-null    int64
 1   Age          891 non-null    float64
 2   Pclass       891 non-null    int64
 3   SibSp        891 non-null    int64
 4   Parch        891 non-null    int64
 5   Fare         891 non-null    float64
 6   Sex          891 non-null    int64
 7   Embarked     891 non-null    int64
dtypes: float64(2), int64(6)
memory usage: 55.8 KB
```

```
[32] sns.heatmap(X.isnull(),yticklabels=False,cbar=False,cmap='viridis')
```

```
<Axes: >
```

Heatmap for Train set to cross-check null values & Training Testing Splitting the model

Titanic Classification.ipynb

File Edit View Insert Runtime Tools Help All changes saved

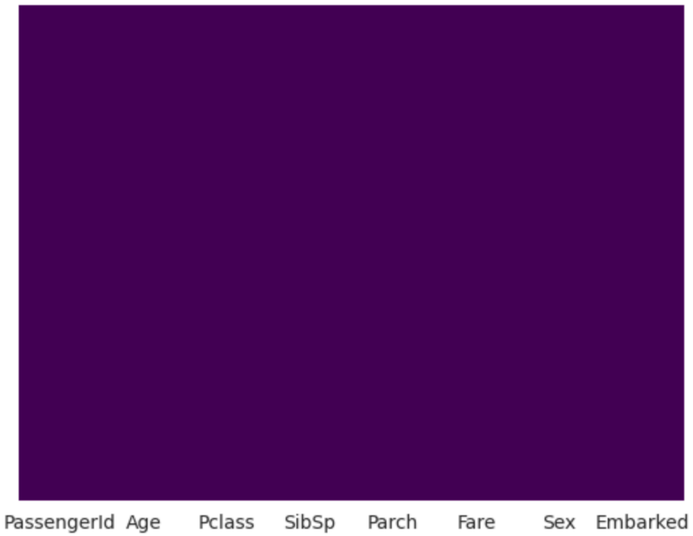
Comment Share

+ Code + Text

RAM Disk Colab AI

```
sns.heatmap(X.isnull(),yticklabels=False,cbar=False,cmap='viridis')
```

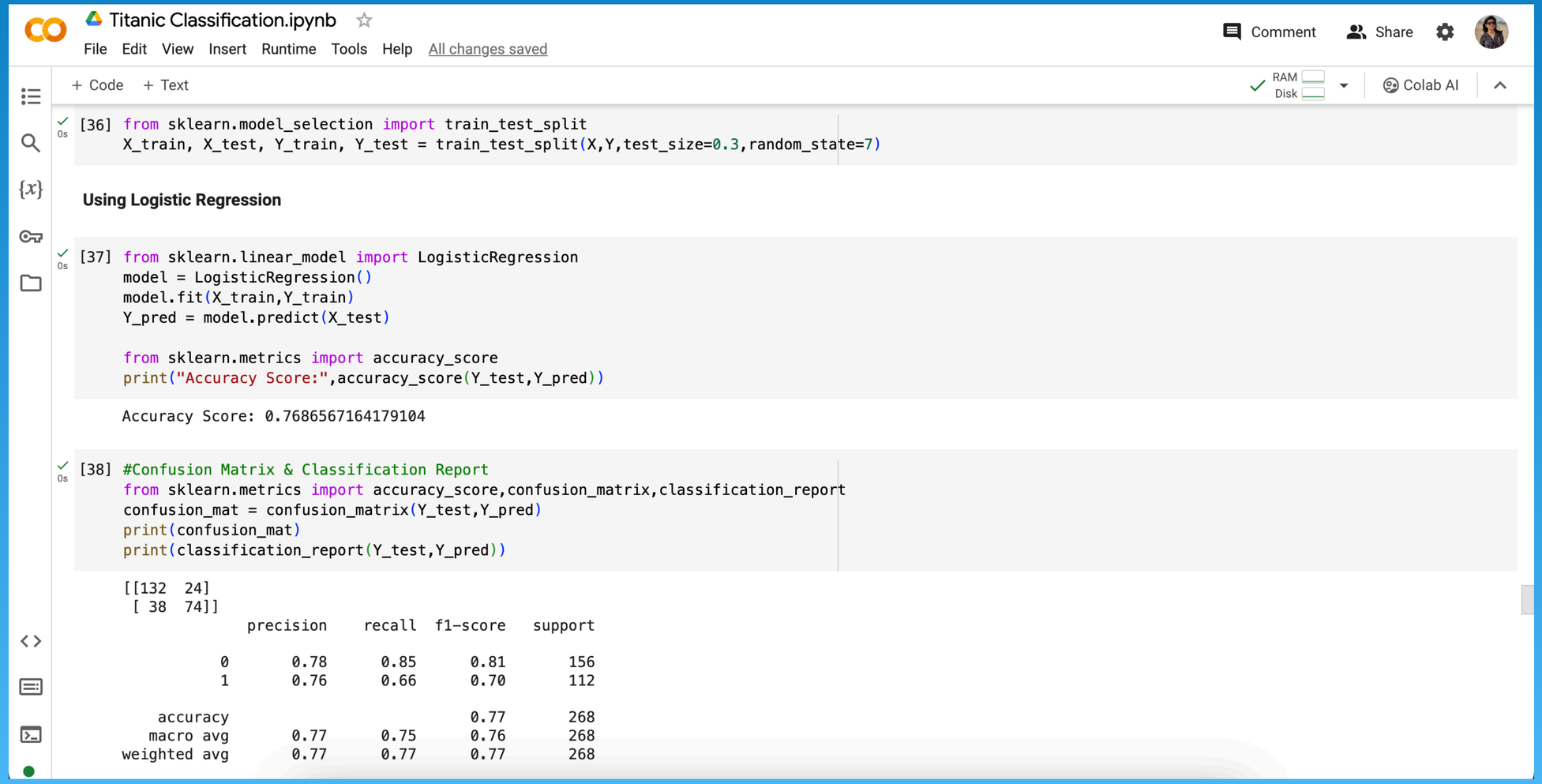
```
<Axes: >
```



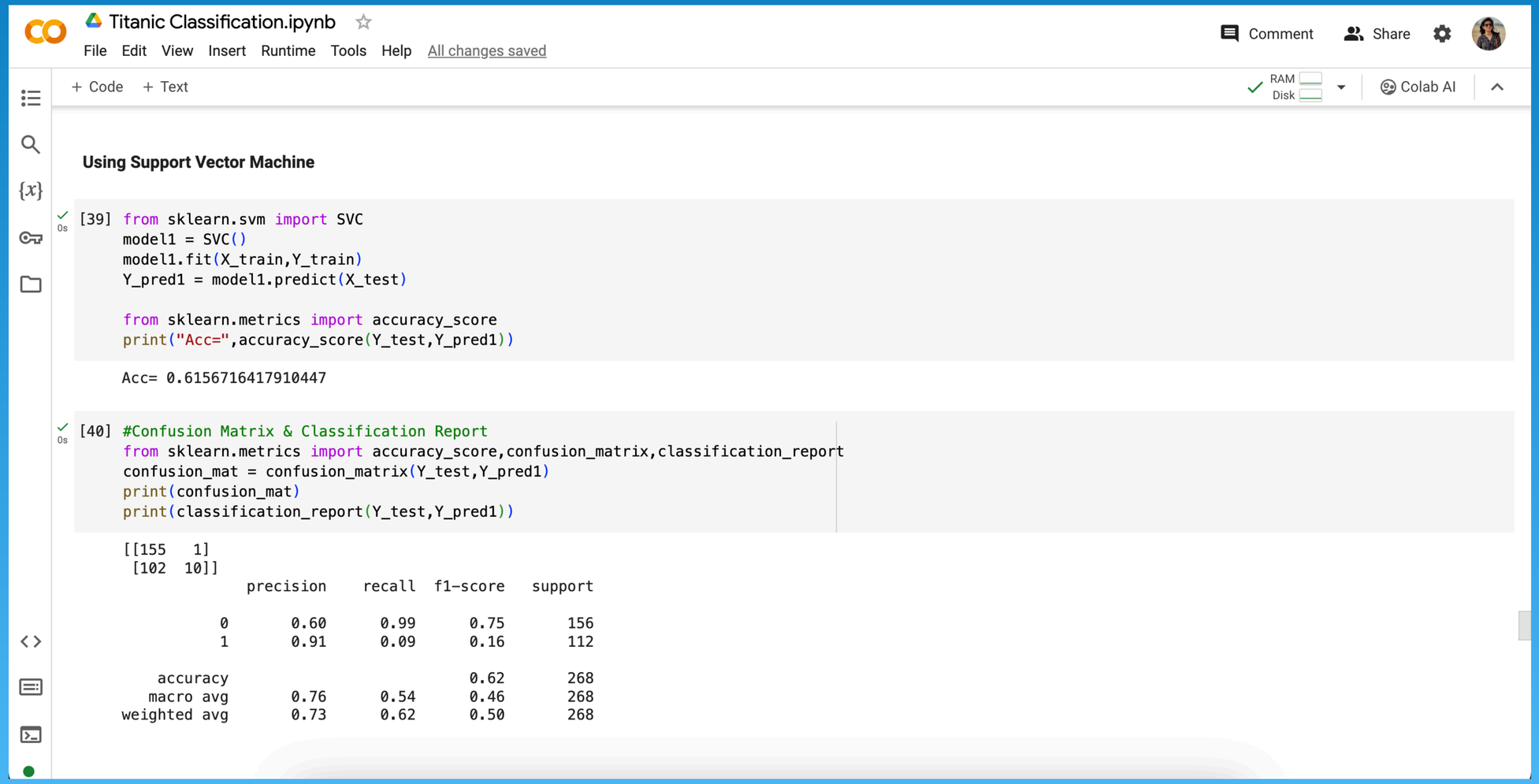
Training Testing and Spliting the model

```
[36] from sklearn.model_selection import train_test_split
     X_train, X_test, Y_train, Y_test = train_test_split(X,Y,test_size=0.3,random_state=7)
```

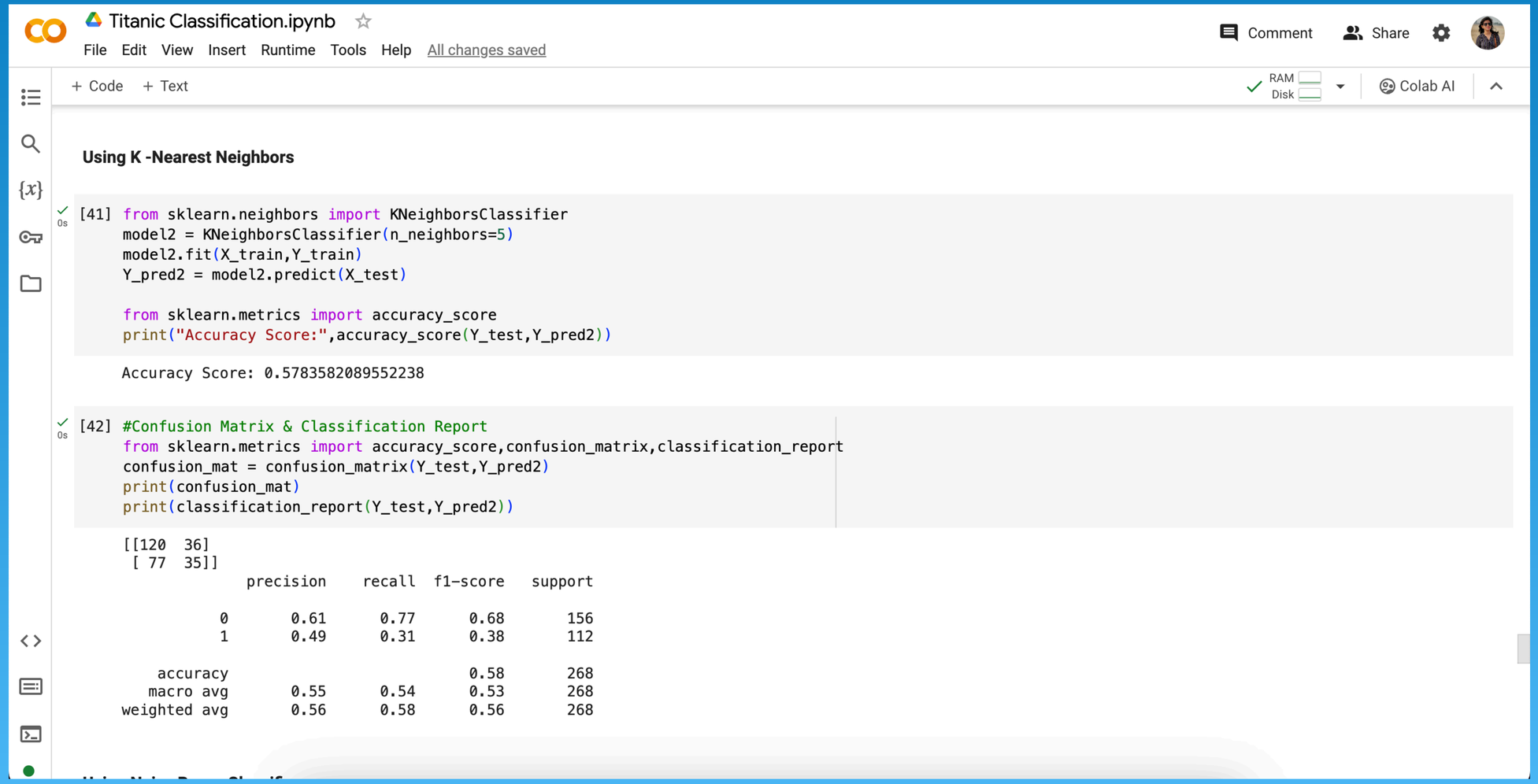
Train-Test Split using Logistic Regression



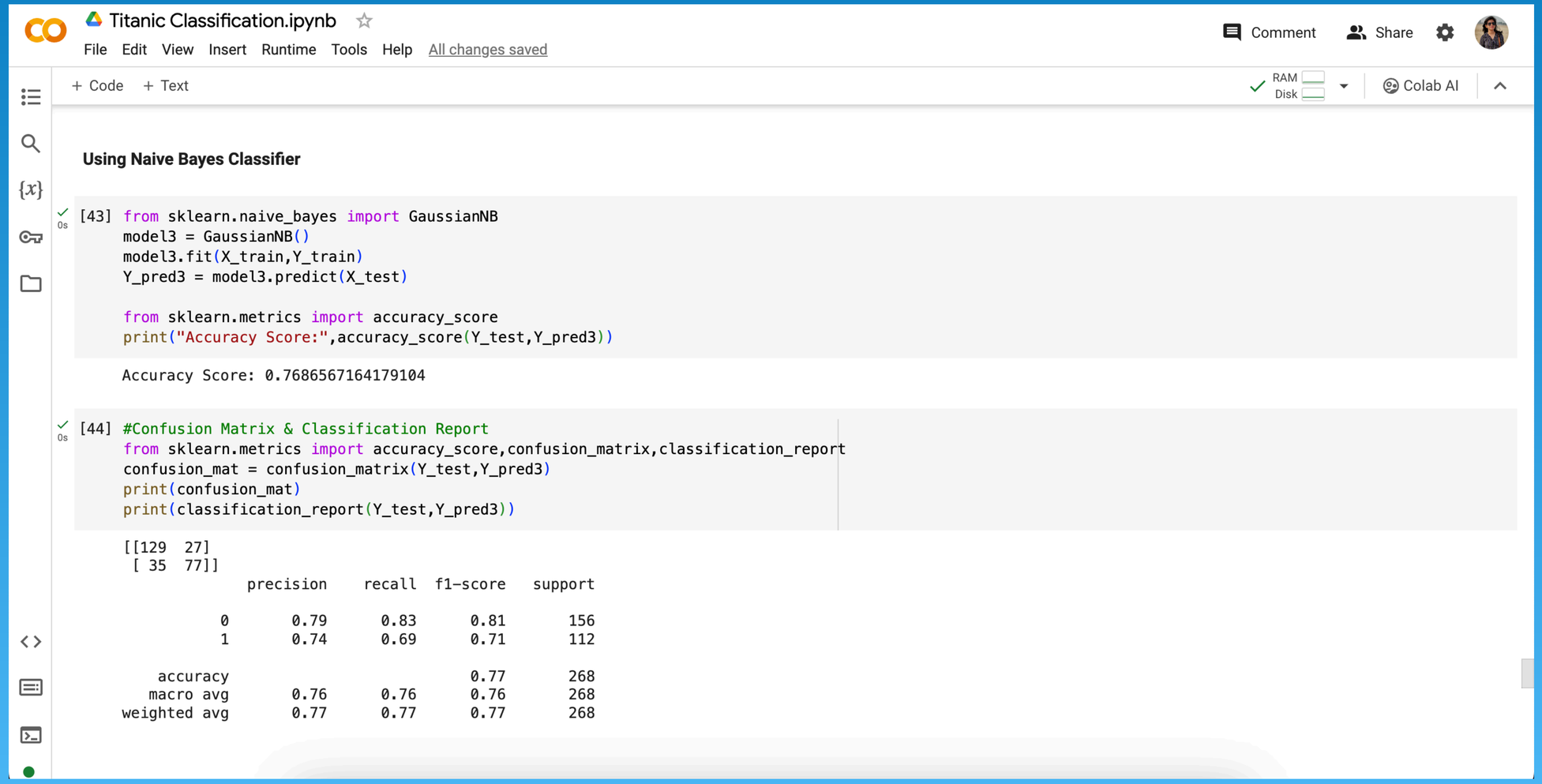
Train-Test Split using Support Vector Machine



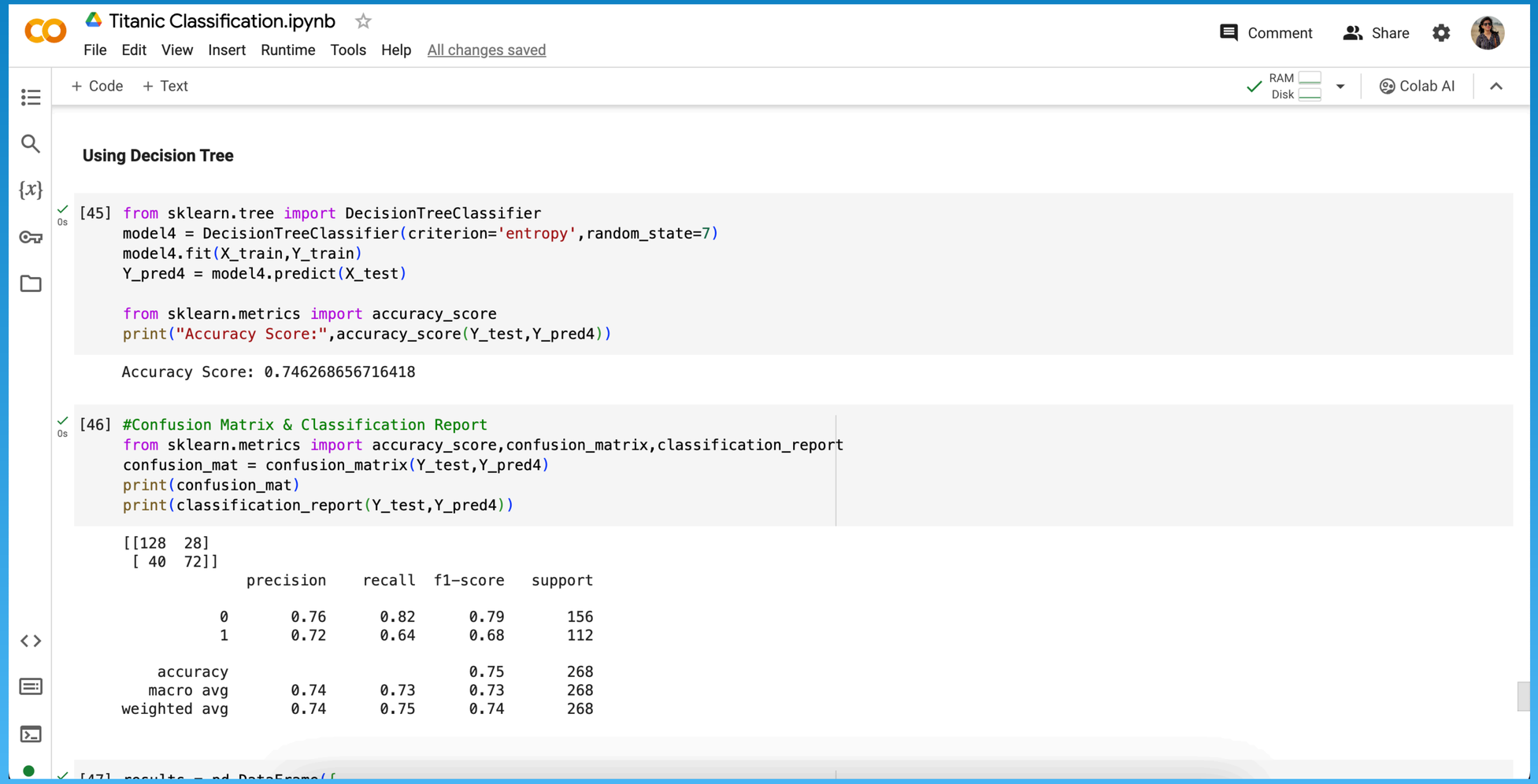
Train-Test Split using K-Nearest Neighbors



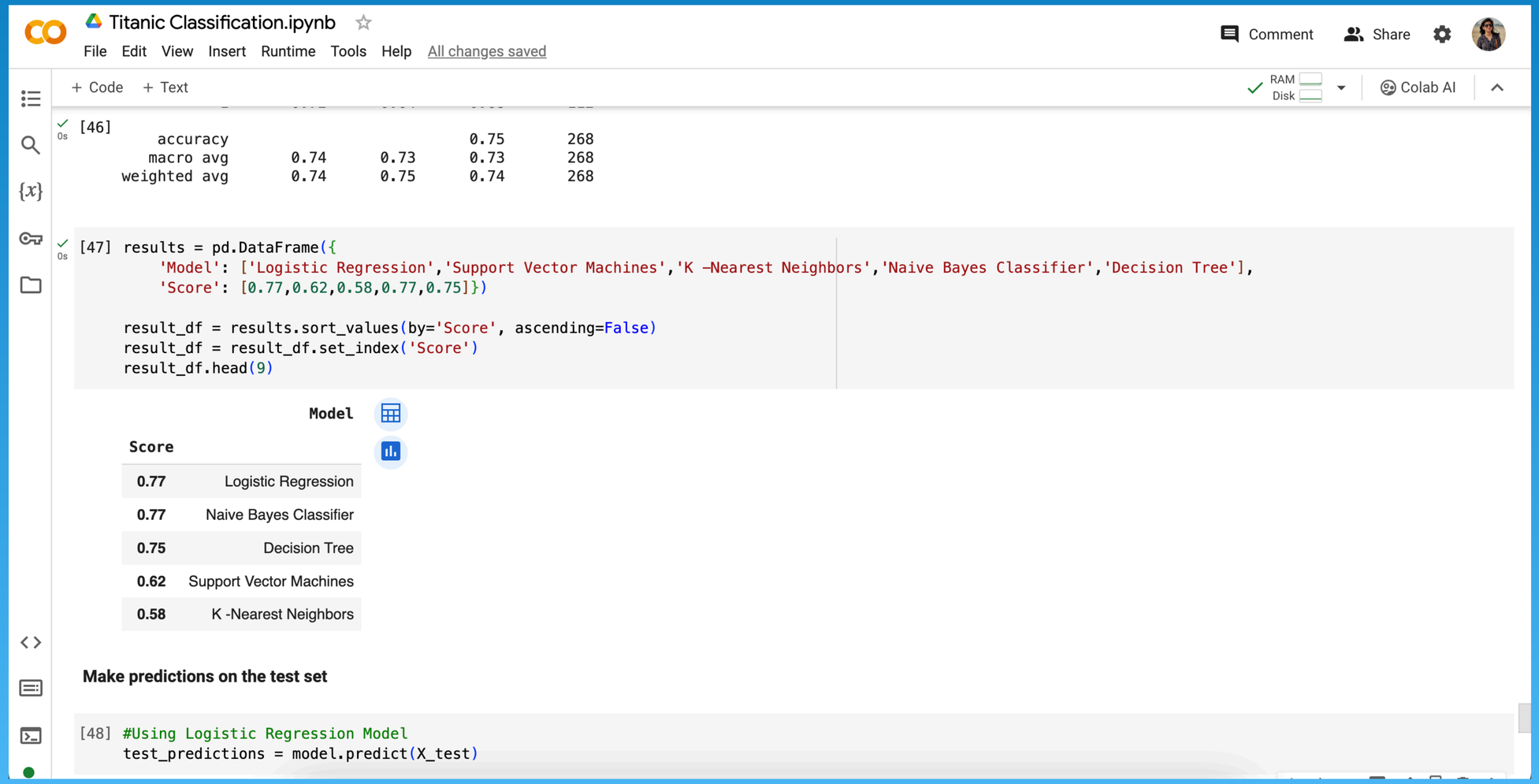
Train-Test Split using Naive Bayes Classifier



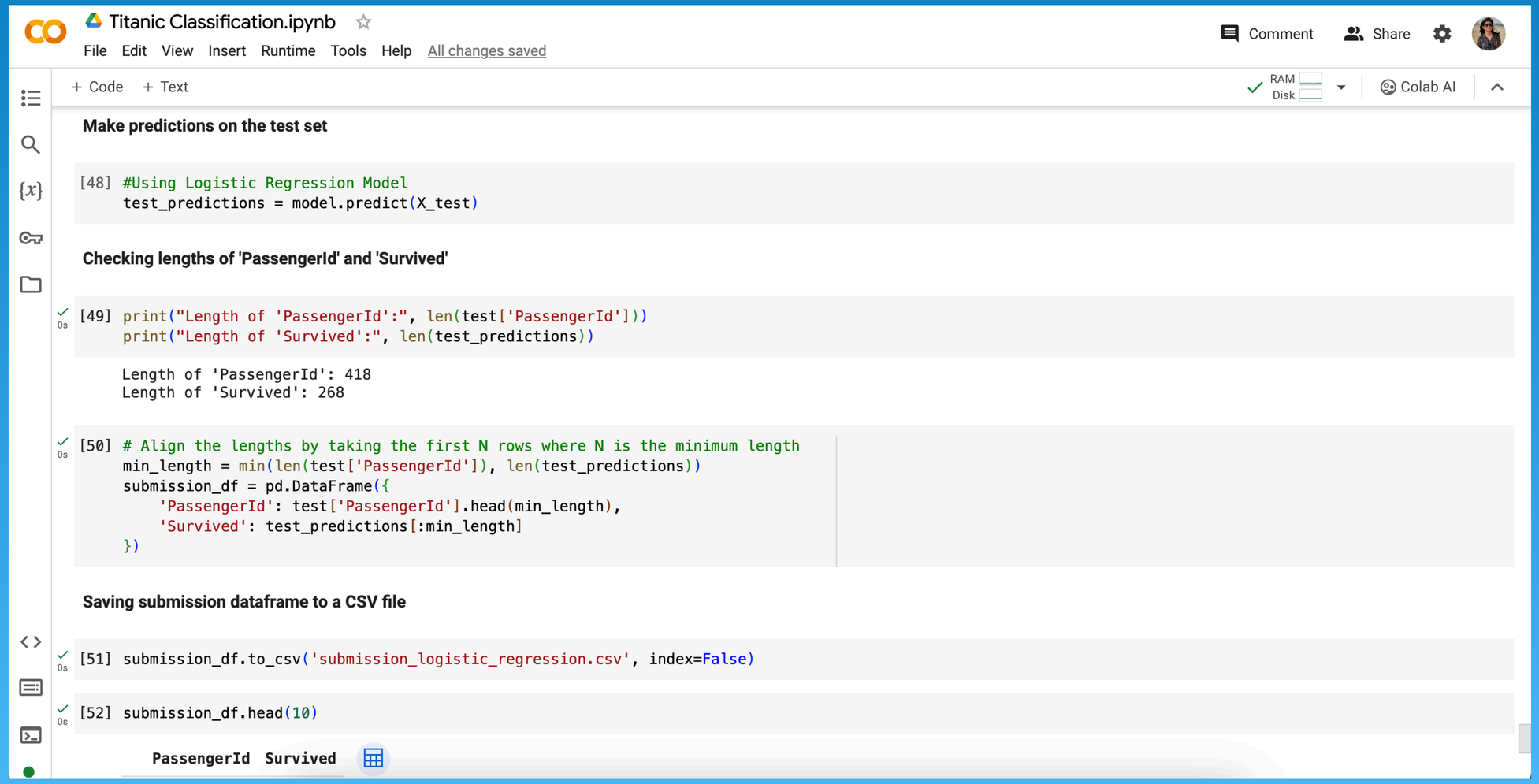
Train-Test Split using Decision Tree



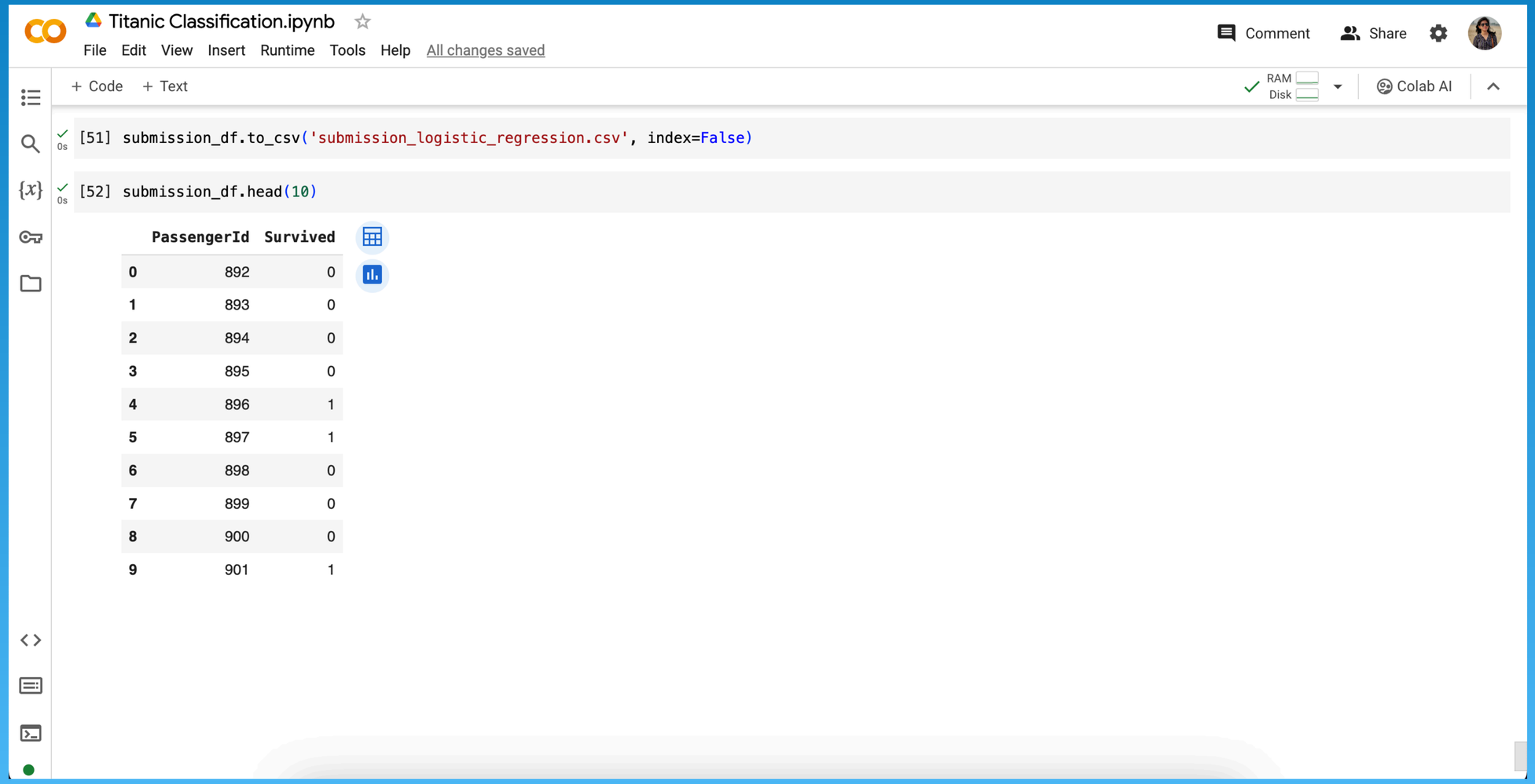
Comparing All used ML Models



Making Predictions on the highest F1 score Model(Logistic Regression) & saving in a csv.file



Fetching the head of newly created dataset



Thank you