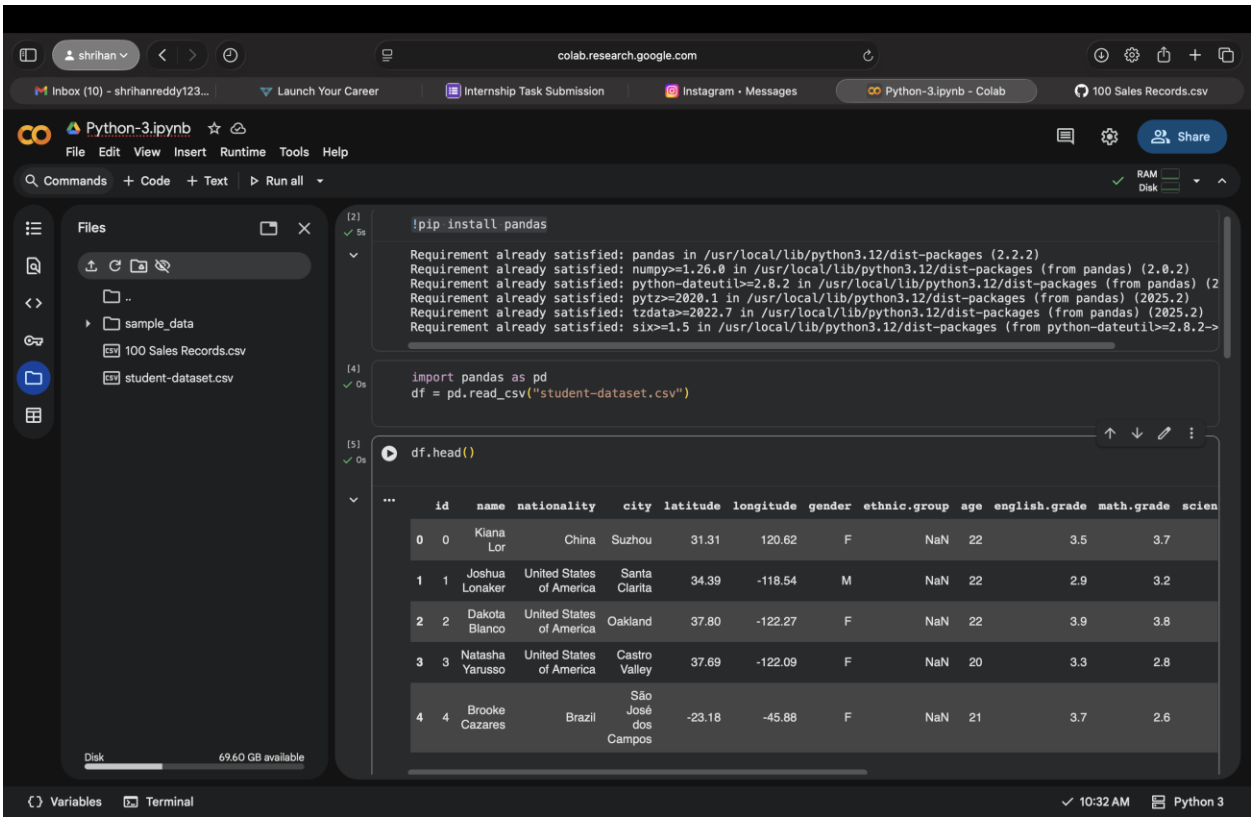
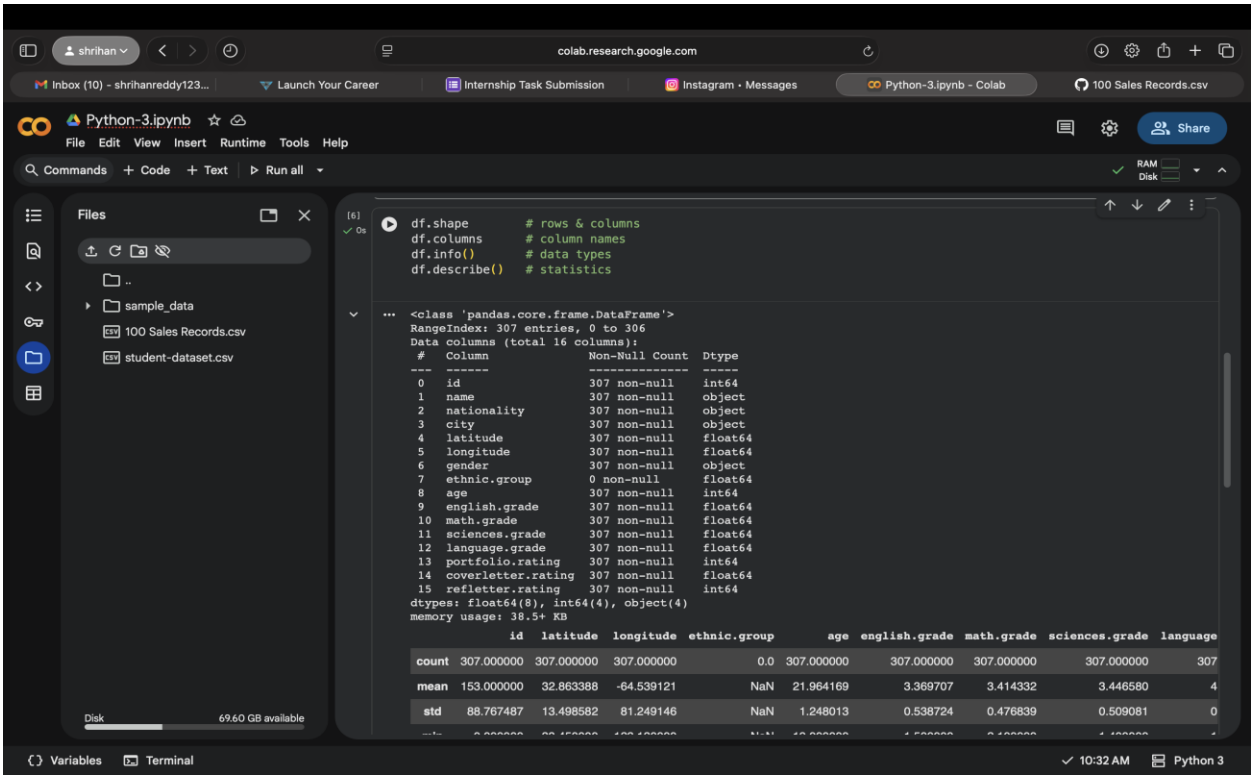
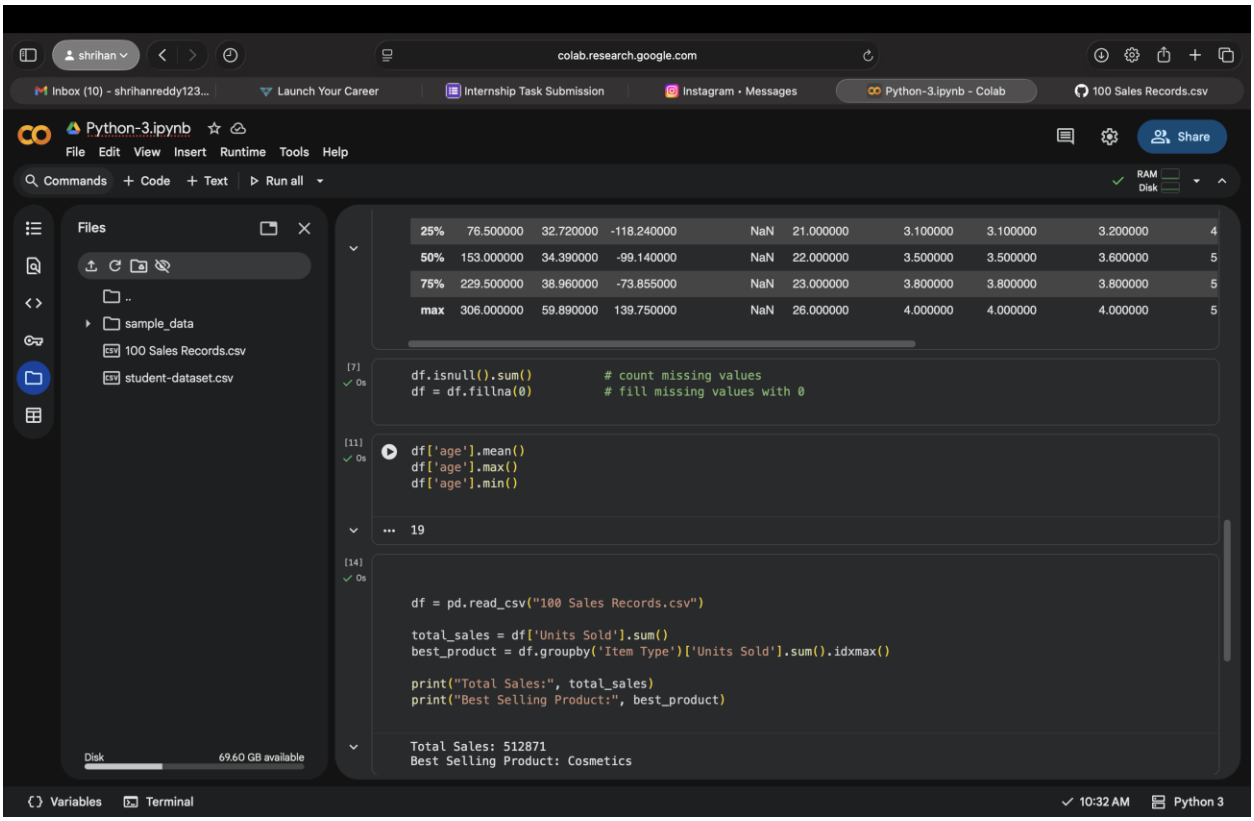

colab.research.google.com
Python-3.ipynb
File Edit View Insert Runtime Tools Help
Share
Commands + Code + Text Run all
Files
sample_data
100 Sales Records.csv
student-dataset.csv
Disk 69.60 GB available
!pip install pandas
Requirement already satisfied: pandas in /usr/local/lib/python3.12/dist-packages (2.2.2)
Requirement already satisfied: numpy>=1.26.0 in /usr/local/lib/python3.12/dist-packages (from pandas) (2.0.2)
Requirement already satisfied: python-dateutil>=2.8.2 in /usr/local/lib/python3.12/dist-packages (from pandas) (2
Requirement already satisfied: pytz>=2020.1 in /usr/local/lib/python3.12/dist-packages (from pandas) (2025.2)
Requirement already satisfied: tzdata>=2022.7 in /usr/local/lib/python3.12/dist-packages (from pandas) (2025.2)
Requirement already satisfied: six>=1.5 in /usr/local/lib/python3.12/dist-packages (from python-dateutil>=2.8.2->
import pandas as pd
df = pd.read_csv("student-dataset.csv")
df.head()
id name nationality city latitude longitude gender ethnic.group age english.grade math.grade scien
0 0 Kiana Lor China Suzhou 31.31 120.62 F NaN 22 3.5 3.7
1 1 Joshua Lonaker United States of America Santa Clarita 34.39 -118.54 M NaN 22 2.9 3.2
2 2 Dakota Blanco United States of America Oakland 37.80 -122.27 F NaN 22 3.9 3.8
3 3 Natasha Yarusso United States of America Castro Valley 37.69 -122.09 F NaN 20 3.3 2.8
4 4 Brooke Cazares Brazil São José dos Campos -23.18 -45.88 F NaN 21 3.7 2.6
Variables Terminal
10:32 AM Python 3


colab.research.google.com
Python-3.ipynb
File Edit View Insert Runtime Tools Help
Share
Commands + Code + Text Run all
Files
sample_data
100 Sales Records.csv
student-dataset.csv
Disk 69.60 GB available
df.shape # rows & columns
df.columns # column names
df.info() # data types
df.describe() # statistics
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 307 entries, 0 to 306
Data columns (total 16 columns):
Column Non-Null Count Dtype
0 id 307 non-null int64
1 name 307 non-null object
2 nationality 307 non-null object
3 city 307 non-null object
4 latitude 307 non-null float64
5 longitude 307 non-null float64
6 gender 307 non-null object
7 ethnic.group 0 non-null float64
8 age 307 non-null int64
9 english.grade 307 non-null float64
10 math.grade 307 non-null float64
11 sciences.grade 307 non-null float64
12 language.grade 307 non-null float64
13 portfolio.rating 307 non-null int64
14 coverletter.rating 307 non-null float64
15 refletter.rating 307 non-null int64
dtypes: float64(8), int64(4), object(4)
memory usage: 38.5+ KB
id latitude longitude ethnic.group age english.grade math.grade sciences.grade language
count 307.000000 307.000000 307.000000 0.0 307.000000 307.000000 307.000000 307.000000 307
mean 153.000000 32.863388 -64.539121 NaN 21.964169 3.369707 3.414332 3.446580 4
std 88.767487 13.498582 81.249146 NaN 1.248013 0.538724 0.476839 0.509081 0
Variables Terminal
10:32 AM Python 3


shrihan
colab.research.google.com
Inbox (10) - shrihanreddy123...
Launch Your Career
Internship Task Submission
Instagram • Messages
Python-3.ipynb - Colab
100 Sales Records.csv
Python-3.ipynb
File Edit View Insert Runtime Tools Help
Share
Commands + Code + Text Run all
RAM
Disk
Files
..
sample_data
100 Sales Records.csv
student-dataset.csv
Disk 69.60 GB available
25% 76.500000 32.720000 -118.240000 NaN 21.000000 3.100000 3.100000 3.200000 4
50% 153.000000 34.390000 -99.140000 NaN 22.000000 3.500000 3.500000 3.600000 5
75% 229.500000 38.960000 -73.855000 NaN 23.000000 3.800000 3.800000 3.800000 5
max 306.000000 59.890000 139.750000 NaN 26.000000 4.000000 4.000000 4.000000 5
[7]
0s
df.isnull().sum() # count missing values
df = df.fillna(0) # fill missing values with 0
[11]
0s
df['age'].mean()
df['age'].max()
df['age'].min()
19
[14]
0s
df = pd.read_csv("100 Sales Records.csv")
total_sales = df['Units Sold'].sum()
best_product = df.groupby('Item Type')['Units Sold'].sum().idxmax()
print("Total Sales:", total_sales)
print("Best Selling Product:", best_product)
Total Sales: 512871
Best Selling Product: Cosmetics
Variables Terminal
10:32 AM Python 3