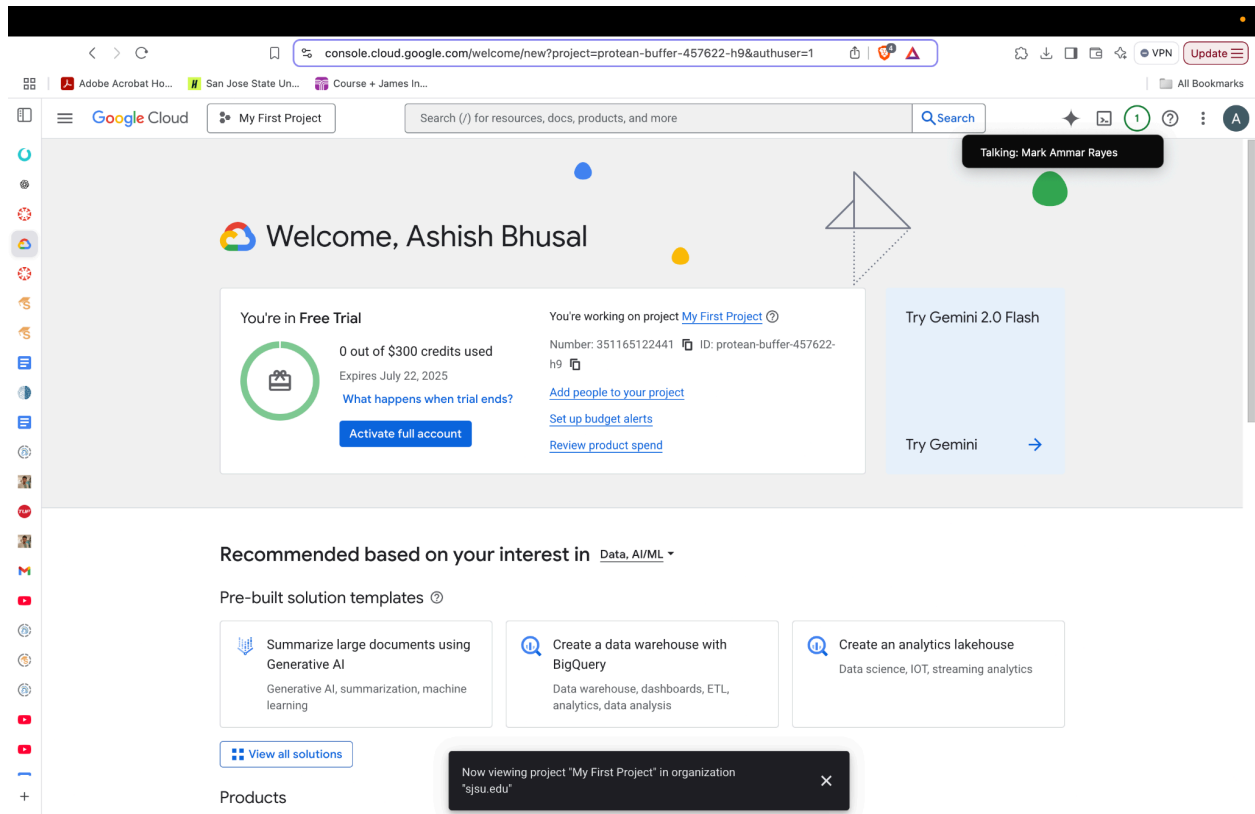


CMPE 252 - AI: Lab 2

Github: https://github.com/bhusalashish/AI_LAB2/

Step 1: Set Up a Free GCP Account



Step 2: Create a Virtual Private Cloud (VPC)

Free trial status: \$300.00 credit and 78 days remaining. Activate your full account to get unlimited access to all of Google Cloud—use any remaining credits, then pay only for what you use. [Dismiss](#) [Activate](#)

Google Cloud My First Project vpc ne Search

VPC Network / VPC networks / Network: sjsu-lab-vpc

VPC networks VPC network details Delete VPC network

IP addresses
Internal ranges
Bring your own IP
Firewall
Routes
VPC network peering
Shared VPC
Serverless VPC access
Packet mirroring
VPC Flow Logs

sjsu-lab-vpc

Overview Subnets Static internal IP addresses Firewalls Firewall endpoints Routes VPC network peering Private services access

Edit

Description
AI and data Engineering lab

Maximum transmission unit
1460

VPC network ULA internal IPv6 range
Disabled

Subnet creation mode
Custom subnets

Dynamic routing mode
Regional

Best path selection mode
Legacy

Tags ⓘ
—

Equivalent REST

Successfully created network "sjsu-lab-vpc".

Step 3: Explore BigQuery

Free trial status: \$300.00 credit and 78 days remaining. Activate your full account to get unlimited access to all of Google Cloud—use any remaining credits, then pay only for what you use.

DismissActivate

Google Cloud

My First Project

BigQuery

Search

Explorer

+ Add data

<

Search BigQuery resources

?

Show starred only

Repositories

Pipelines

External connections

america_health_r...

austin_311

austin_bikeshare

bikeshare_stat...

bikeshare_trips

austin_crime

austin_incidents

austin_waste

baseball

bbc_news

Repository

Preview

>

No repository selected

Select a repository and a workspace to view its content.

View repositories

bikeshare_trips

Query

Open in

Share

Copy

Snapshot

Delete

Export

Refresh

Schema

Details

Preview

Table Explorer

Preview

Insights

Lineage

Data Profile

Data Quality

Filter

Enter property name or value

?

| <input type="checkbox"/> | Field name | Type | Mode | Key | Collation | Default Value | Policy Tags | Description |
|--------------------------|--------------------|-----------|----------|-----|-----------|---------------|-------------|-------------------------|
| <input type="checkbox"/> | trip_id | STRING | NULLABLE | - | - | - | - | Numeric ID of bike trip |
| <input type="checkbox"/> | subscriber_type | STRING | NULLABLE | - | - | - | - | Type of the Subscriber |
| <input type="checkbox"/> | bike_id | STRING | NULLABLE | - | - | - | - | ID of bike used |
| <input type="checkbox"/> | bike_type | STRING | NULLABLE | - | - | - | - | Type of bike used |
| <input type="checkbox"/> | start_time | TIMESTAMP | NULLABLE | - | - | - | - | Start timestamp of trip |
| <input type="checkbox"/> | start_station_id | INTEGER | NULLABLE | - | - | - | - | Numeric referen... |
| <input type="checkbox"/> | start_station_name | STRING | NULLABLE | - | - | - | - | Station name for... |
| <input type="checkbox"/> | end_station_id | STRING | NULLABLE | - | - | - | - | Numeric referen... |
| <input type="checkbox"/> | end_station_name | STRING | NULLABLE | - | - | - | - | Station name for... |
| <input type="checkbox"/> | duration_minutes | INTEGER | NULLABLE | - | - | - | - | Time of trip in minutes |

Edit schema

View row access policies

Job history

'bigquery-public-data' was starred.

Refresh

Free trial status: \$300.00 credit and 78 days remaining. Activate your full account to get unlimited access to all of Google Cloud—use any remaining credits, then pay only for what you use.

Dismiss

Activate

Google Cloud

My First Project

BigQuery

Search

Explorer

+ Add data

<

Search BigQuery resources

?

Show starred only

Repositories

Pipelines

External connections

america_health_r...

austin_311

austin_bikeshare

bikeshare_stat...

bikeshare_trips

austin_crime

austin_incidents

austin_waste

baseball

bbc_news

Repository

Preview

>

No repository selected

Select a repository and a workspace to view its content.

View repositories

bikeshare_trips

Query

Open

Share

Copy

Snapshot

Delete

Export

Refresh

Schema

Details

Preview

Table Explorer

Preview

Insights

Lineage

Data Profile

Data Quality

| Row | trip_id | subscriber_type | bike_id | bike_type | start_time | sta |
|-----|----------|-------------------------------|---------|-----------|-------------------------|-----|
| 1 | 26599763 | Pay-as-you-ride | 21707 | electric | 2022-05-06 14:19:39 UTC | |
| 2 | 26742903 | 3-Day Weekender | 17460 | electric | 2022-05-23 16:24:46 UTC | |
| 3 | 26599923 | Pay-as-you-ride | 19453 | electric | 2022-05-06 14:37:41 UTC | |
| 4 | 26701683 | Local31 | 21772 | electric | 2022-05-17 22:50:29 UTC | |
| 5 | 26788653 | Pay-as-you-ride | 21740 | electric | 2022-05-29 19:41:40 UTC | |
| 6 | 26742937 | 3-Day Weekender | 17460 | electric | 2022-05-23 16:28:34 UTC | |
| 7 | 26742936 | 3-Day Weekender | 18726 | electric | 2022-05-23 16:28:33 UTC | |
| 8 | 26648461 | Student Membership | 22599 | electric | 2022-05-12 09:38:57 UTC | |
| 9 | 26599753 | Pay-as-you-ride | 21691 | electric | 2022-05-06 14:18:39 UTC | |
| 10 | 26124027 | Local31 | 21419 | electric | 2022-03-06 21:51:17 UTC | |
| 11 | 26254020 | Single Trip (Pay-as-you-ride) | 226 | classic | 2022-03-26 22:47:08 UTC | |
| 12 | 26254013 | Single Trip (Pay-as-you-ride) | 1703 | classic | 2022-03-26 22:46:37 UTC | |
| 13 | 26124026 | Pay-as-you-ride | 21740 | electric | 2022-03-06 21:50:56 UTC | |
| 14 | 26163419 | Pay-as-you-ride | 19221 | electric | 2022-03-14 21:31:48 UTC | |
| 15 | 26168832 | Explorer | 21428 | electric | 2022-03-15 17:04:59 UTC | |
| 16 | 26124031 | Pay-as-you-ride | 21516 | electric | 2022-03-06 21:53:49 UTC | |
| 17 | 26253986 | Single Trip (Pay-as-you-ride) | 0190 | classic | 2022-03-26 22:41:56 UTC | |
| 18 | 26363471 | 3-Day Weekender | 19285 | electric | 2022-04-09 18:33:15 UTC | |
| 19 | 26423542 | Student Membership | 1641 | classic | 2022-04-16 19:42:03 UTC | |
| 20 | 26466475 | Local365 | 12727 | classic | 2022-04-21 20:43:06 UTC | |
| 21 | 27313573 | 24 Hour Walk Up Pass | 21454 | electric | 2022-07-27 19:02:38 UTC | |

Results per page: 50

1 - 50 of 2271152

<

>

|

Job history

Refresh

^

The screenshot displays the Google Cloud BigQuery console interface. At the top, a navigation bar includes the Google Cloud logo, project name 'My First Project', and a search bar. Below this, a sidebar on the left contains an 'Explorer' panel with a search bar and a list of resources including 'Repositories', 'Pipelines', 'External connections', 'america_health_r...', 'austin_311', 'austin_bikeshare', 'bikeshare_stat...', 'bikeshare_trips', 'austin_crime', 'austin_incidents', 'austin_waste', 'baseball', and 'bbc_news'. A 'Repository' section below the sidebar shows 'No repository selected' and a 'View repositories' button. The main area features a 'Query editor' with a SQL query:

```
1 SELECT bike_id, duration_minutes
2 FROM `bigquery-public-data.austin_bikeshare.bikeshare_trips`
3 LIMIT 10
```

. Below the editor, a 'Query completed' message is shown. The 'Query results' section displays a table with columns 'bike_id' and 'duration_minutes'. The table contains 10 rows of data. At the bottom, a 'Job history' section is visible with a 'Refresh' button.

| Row | bike_id | duration_minutes |
|-----|---------|------------------|
| 1 | 198BB | 190 |
| 2 | 133G | 44 |
| 3 | 589 | 7 |
| 4 | 302 | 2 |
| 5 | 391 | 3 |
| 6 | 956 | 9 |
| 7 | 106G | 12 |
| 8 | 198BB | 10 |
| 9 | 936 | 11 |
| 10 | 113 | 10 |

Step 4: Explore AutoML (Vertex AI)

Part A: Upload CSV to Cloud Storage

Free trial status: \$300.00 credit and 78 days remaining. Activate your full account to get unlimited access to all of Google Cloud—use any remaining credits, then pay only for what you use.

Dismiss

Activate

Google Cloud

My First Project

Search (/) for resources, docs, products, and more

Search

2

Cloud Storage

Bucket details

Go to path

Refresh

Overview

Buckets

Monitoring

Settings

Storage Intelligence

Insights datasets

Configuration

Marketplace

Release Notes

018320627-bucket

Location

Storage class

Public access

Protection

us-west1 (Oregon)

Standard

Not public

Soft Delete

Objects

Configuration

Permissions

Protection

Lifecycle

Observability

New

Inventory Reports

Operations

Folder browser

018320627-bucket

Create folder

Upload

Transfer data

Other services

Learn

Filter by name prefix only

Filter

Filter objects and folders

Show

Live objects only

| <input type="checkbox"/> | Name | Size | Type | Created | Storage class | |
|--------------------------|-----------------|--------|----------|-------------------------|---------------|--|
| <input type="checkbox"/> | Customers 2.csv | 1.8 MB | text/csv | May 5, 2025, 2:25:42 PM | Standard | |

Part B: Open and Use Notebook

The screenshot shows the Google Cloud Vertex AI console interface. The browser address bar displays the URL: `console.cloud.google.com/vertex-ai/locations/us-west1/datasets/6671194442554146816/im...`. The page title is "sample-auto-ml-lab".

Left Sidebar (Tools):

- Dashboard
- Model Garden
- Pipelines
- Notebooks
- Vertex AI Studio
- Agent builder
- Data
- Feature Store
- Datasets (highlighted)
- Model development
 - Training
 - Experiments
 - Metadata
 - Ray on Vertex AI
- Deploy and use
 - Model Registry
 - Marketplace

Main Content Area:

Source | Analyze | Lineage

Optimized for the best results.

[View data guide](#)

Select a data source

- CSV file: Can be uploaded from your computer or on Cloud Storage. [Learn more](#)
- BigQuery: Select a table or view from BigQuery. [Learn more](#)

☐ Upload CSV files from your computer

☒ Select CSV files from Cloud Storage

☐ Select a table or view from BigQuery

Select CSV files from Cloud Storage

Enter the Cloud Storage path to one or more CSV files. Data from multiple files will be referenced as one dataset.

Import file path * [Browse](#)

[Add another file](#)

What happens next?

The selected CSV file will be associated with your dataset. Making changes to the referenced CSV file will affect the dataset.

[Continue](#)

Right Panel (Model Examples):

Two model cards are shown:

- House icon: \$625,000
- House icon: \$975,000

You can build two model types with tabular data. The model type is automatically chosen based on the data type of your target column.

- Regression models predict a numeric value. For example, predicting home prices or consumer spending.
- Classification models predict a category from a fixed number of categories. Examples include predicting whether an email is spam or not, or classes a student might be interested in attending.

Free trial status: \$300.00 credit and 78 days remaining. Activate your full account to get unlimited access to all of Google Cloud—use any remaining credits, then pay only for what you use. [Dismiss](#) [Activate](#)

Google Cloud My First Project Search (/) for resources, docs, products, and more Search

Vertex AI sample-auto-ml-lab

Tools

- Dashboard
- Model Garden
- Pipelines
- Notebooks
- Vertex AI Studio
- Agent builder
- Data
- Feature Store
- Datasets

Model development

- Training
- Experiments
- Metadata
- Ray on Vertex AI

Deploy and use

- Model Registry
- Online prediction
- Marketplace

Source Analyze Lineage

Analyze

Properties

| | |
|---------------------|--|
| Created | May 05, 2025 2:33 PM |
| Dataset format | CSV |
| Dataset location(s) | gs://018320627-buc-t/Customers 2.csv |
| Encryption type | Google-managed |

Summary

Total columns: 1

Total rows: -

Generating statistics...

Filter Enter property name or value

Column name ↑

Index;Customer Id;First Name;Last Name;Company;City;Country;Phone 1;Phone 2;Email;Subscription Date;Website

Dark mode Public Preview is now available in certain products

To turn it on, open settings, select preferences, and choose dark mode

[Go to settings](#)

Training jobs and models

Use this dataset and annotation set to train a new machine learning model with AutoML or custom code. Selecting **AutoML** on Pipelines will create a Run on Vertex AI Pipelines. Run information will be found on the [Runs](#) tab under Pipelines.

[Train new model](#)

Step 5: Explore AI Platform

Free trial status: \$300.00 credit and 78 days remaining. Activate your full account to get unlimited access to all of Google Cloud—use any remaining credits, then pay only for what you use. [Dismiss](#) [Activate](#)

Google Cloud My First Project AI Search

Vertex AI / Workbench / Instances

Tools

- Dashboard
- Model Garden
- Pipelines

Notebooks

- Colab Enterprise
- Workbench

Vertex AI Studio

Agent builder

Data

- Feature Store
- Datasets

Model development

- Training
- Experiments
- Metadata
- Ray on Vertex AI

Deploy and use

- Marketplace

Workbench [Create New](#) [Refresh](#) [Learn](#)

[Instances](#) Executions Schedules

View: [Instances](#) User-managed Notebooks Managed Notebooks

JupyterLab 4 is now available in Vertex AI Workbench. [Dismiss](#)

Workbench Instances have JupyterLab 3 pre-installed and are configured with GPU-enabled machine learning frameworks. [Learn more](#)

Filter

| | Instance name | Zone | Auto upgrade | Version | Machine Type | GPUs | Owner | Created |
|--------------------------|--|------------|--------------|---------|--|------|--|-------------------------|
| <input type="checkbox"/> | instance-20250505-144443 | us-west1-a | — | — | Efficient Instance: 4 vCPUs, 16 GB RAM | None | 351165122441-compute@developer.gserviceaccount.com | May 5, 2025, 2:46:19 PM |

Read and Analyze customers.csv

6610735966799491-dot-us-west1.notebooks.googleusercontent.com/lab/tree/AL_Lab...

Adobe Acrobat Ho... San Jose State Un... Course + James In...

instance-20250505-144443

e2-standard-4

File Edit View Run Kernel Git Tabs Settings Help

Launcher

notebook_template.ipynb AL_Lab2.ipynb

Python 3 (ipykernel) (Local)

Filter files by name

/

Name

AL_Lab2.ipynb notebook_template...

Install Required Libraries

+ 1 cell hidden

Import Libraries

+ 1 cell hidden

Set Up GCS Access

[11]:

BUCKET_NAME = '018320627-bucket'
FILE_NAME = 'Customers.csv'

client = storage.Client()
bucket = client.bucket(BUCKET_NAME)
blob = bucket.blob(FILE_NAME)

Read content as string
data = blob.download_as_text()
df = pd.read_csv(StringIO(data), sep=';')
df.head()

print(data)

[11]:

| | Index | Customer Id | First Name | Last Name | Company | City | Country | Phone 1 | Phone 2 | Email | Subscription Date | Website |
|---|-------|-----------------|------------|-----------|--------------------------|------------------|--------------------------|--------------|-------------------|-------------------------------|-------------------|-----------------------------|
| 0 | 1 | EB54EF1154C3A78 | Heather | Callahan | Mosley-David | Lake Jeffborough | Norway | 043-797-5229 | 915.112.1727 | urangel@espinoza-francis.net | 2020-08-26 | http://www.escobar.org |
| 1 | 2 | 10dAcafEBbA5FcA | Kristina | Ferrell | Horn, Shepard and Watson | Aaronville | Andorra | 932-062-1802 | (209)72-7124x3651 | xreese@hall-donovan.com | 2020-04-27 | https://tyler-pugh.info |
| 2 | 3 | 67DAB15Ebe4BE4a | Briana | Andersen | Irwin-Oneal | East Jordan | Nepal | 8352752061 | (567)135-1918 | haleybraun@blevins-sexton.com | 2022-03-22 | https://www.mackbell.net |
| 3 | 4 | 6d350C5E5eDB4EE | Patty | Ponce | Richardson Group | East Kristintown | Northern Mariana Islands | 302.398.3833 | 196-189-7767x770 | hohailey@anthony.com | 2020-07-02 | https://delacruzfreeman.org |

Simple 0 2 Python 3 (ipykernel) (Local) Idle Mode: Command Ln 1, Col 19 AL_Lab2.ipynb 1

The screenshot shows a Jupyter Notebook titled "Do Some Basic Analysis". The code in cell [18] performs the following actions:

- Prints basic information about the DataFrame.
- Prints summary statistics.
- Prints the distribution of genders (if it exists).
- Prints the average income (if it exists).

The output of the code is as follows:

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 10000 entries, 0 to 9999
Data columns (total 15 columns):
#   Column                Non-Null Count  Dtype
---  ---
0   Index                  10000 non-null  int64
1   Customer Id            10000 non-null  object
2   First Name              10000 non-null  object
3   Last Name               10000 non-null  object
4   Company                10000 non-null  object
5   City                   10000 non-null  object
6   Country                 10000 non-null  object
7   Phone 1                 10000 non-null  object
8   Phone 2                 10000 non-null  object
9   Email                  10000 non-null  object
10  Subscription Date       10000 non-null  object
11  Website                 10000 non-null  object
12  Age                     10000 non-null  int64
13  Annual Income (k$)      10000 non-null  int64
14  Spending Score (1-100)  10000 non-null  int64
dtypes: int64(4), object(11)
memory usage: 1.1+ MB
Average income: 60.56
```

Short answer Questions

What is the average value of a numeric column (e.g., income or age)?

=> The average income is 60.56 and average age is 38.85

```
[21]: print("Average income:", df['Annual Income (k$)'].mean())
      print("Average age:", df['Age'].mean())
```

```
Average income: 60.56
Average age: 38.85
```

How many unique values did you observe in the Gender or Category column?

=> There is no such column as 'Gender' or 'Category' in the given dataset

```
[24]: print(df.columns)
if not "Gender" in df.columns:
    print("'Gender' column does not exists")

Index(['Index', 'Customer Id', 'First Name', 'Last Name', 'Company', 'City',
      'Country', 'Phone 1', 'Phone 2', 'Email', 'Subscription Date',
      'Website', 'Age', 'Annual Income (k$)', 'Spending Score (1-100)'],
      dtype='object')
'Gender' column does not exists
```

Were there any missing (NaN) values? Which column had the most?

=> There is no any column with missing values.

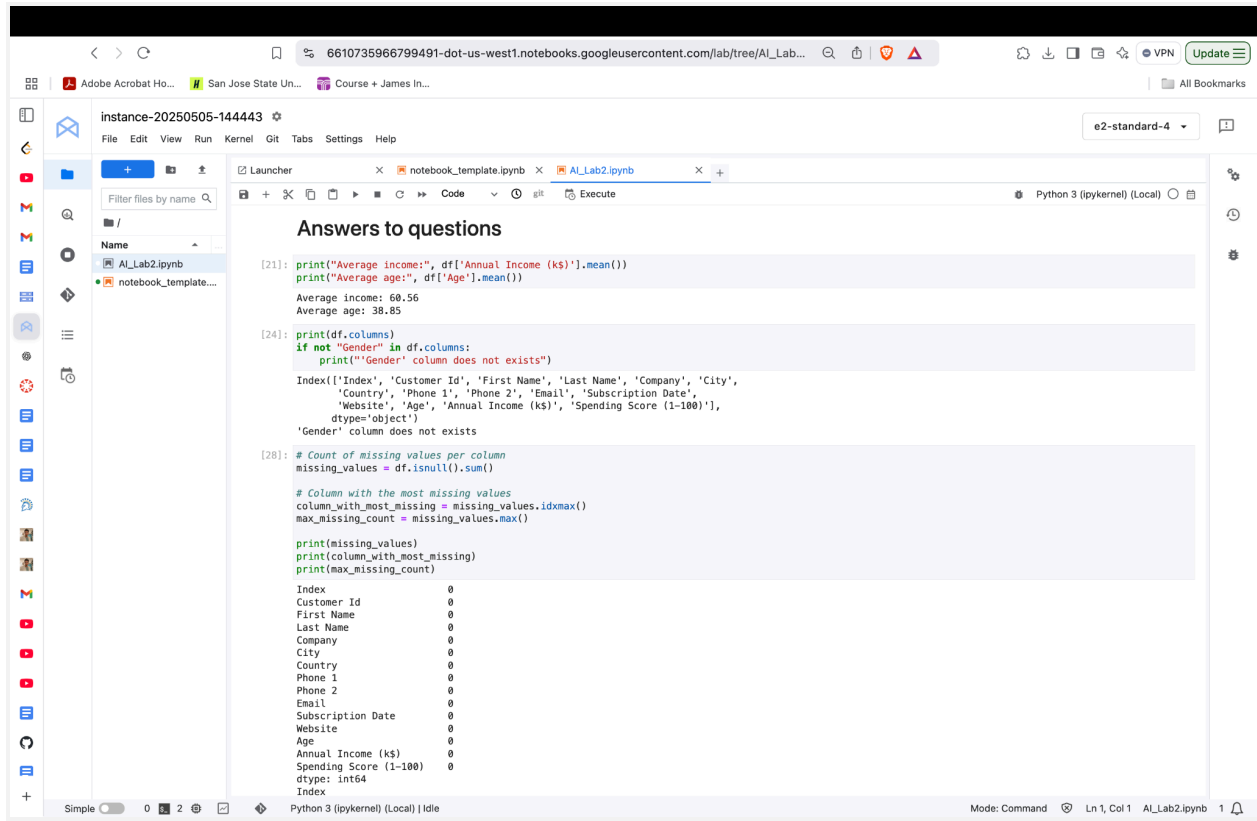
```
[28]: # Count of missing values per column
missing_values = df.isnull().sum()

# Column with the most missing values
column_with_most_missing = missing_values.idxmax()
max_missing_count = missing_values.max()

print(missing_values)
print(column_with_most_missing)
print(max_missing_count)

Index          0
Customer Id    0
First Name     0
Last Name      0
Company        0
City           0
Country        0
Phone 1        0
Phone 2        0
Email          0
Subscription Date 0
Website        0
Age            0
Annual Income (k$) 0
Spending Score (1-100) 0
dtype: int64
Index
0
```

Answers



The screenshot shows a Jupyter Notebook interface with the following components:

- Top Bar:** Includes a browser address bar with the URL `6610735966799491-dot-us-west1.notebooks.googleusercontent.com/lab/tree/AI_Lab...` and a toolbar with icons for file operations, a VPN button, and an 'Update' button.
- Left Sidebar:** Displays the file explorer for the instance `instance-20250505-144443`. It shows a file named `AI_Lab2.ipynb` and a folder named `notebook_template...`.
- Main Area:** Contains the code editor for `AI_Lab2.ipynb`. The code is titled "Answers to questions" and includes the following Python code:

```
[21]: print("Average income:", df['Annual Income (k$)'].mean())
      print("Average age:", df['Age'].mean())
      Average income: 60.56
      Average age: 38.85

[24]: print(df.columns)
      if not "Gender" in df.columns:
          print("'Gender' column does not exists")

      Index(['Index', 'Customer Id', 'First Name', 'Last Name', 'Company', 'City',
            'Country', 'Phone 1', 'Phone 2', 'Email', 'Subscription Date',
            'Website', 'Age', 'Annual Income (k$)', 'Spending Score (1-100)'],
            dtype=object)
      'Gender' column does not exists

[28]: # Count of missing values per column
      missing_values = df.isnull().sum()

      # Column with the most missing values
      column_with_most_missing = missing_values.idxmax()
      max_missing_count = missing_values.max()

      print(missing_values)
      print(column_with_most_missing)
      print(max_missing_count)
```

The output of the code is as follows:

```
Index
Customer Id
First Name
Last Name
Company
City
Country
Phone 1
Phone 2
Email
Subscription Date
Website
Age
Annual Income (k$)
Spending Score (1-100)
dtype: int64
Index
```

- Bottom Bar:** Shows the status of the kernel as `Python 3 (ipykernel) (Local) | Idle` and the current mode as `Simple`. It also displays the cursor position as `Ln 1, Col 1` and the file name `AI_Lab2.ipynb`.

Link for the notebook

https://github.com/bhusalashish/AI_LAB2/blob/main/AI_Lab2.ipynb