

nycflights13

Data Science



Links

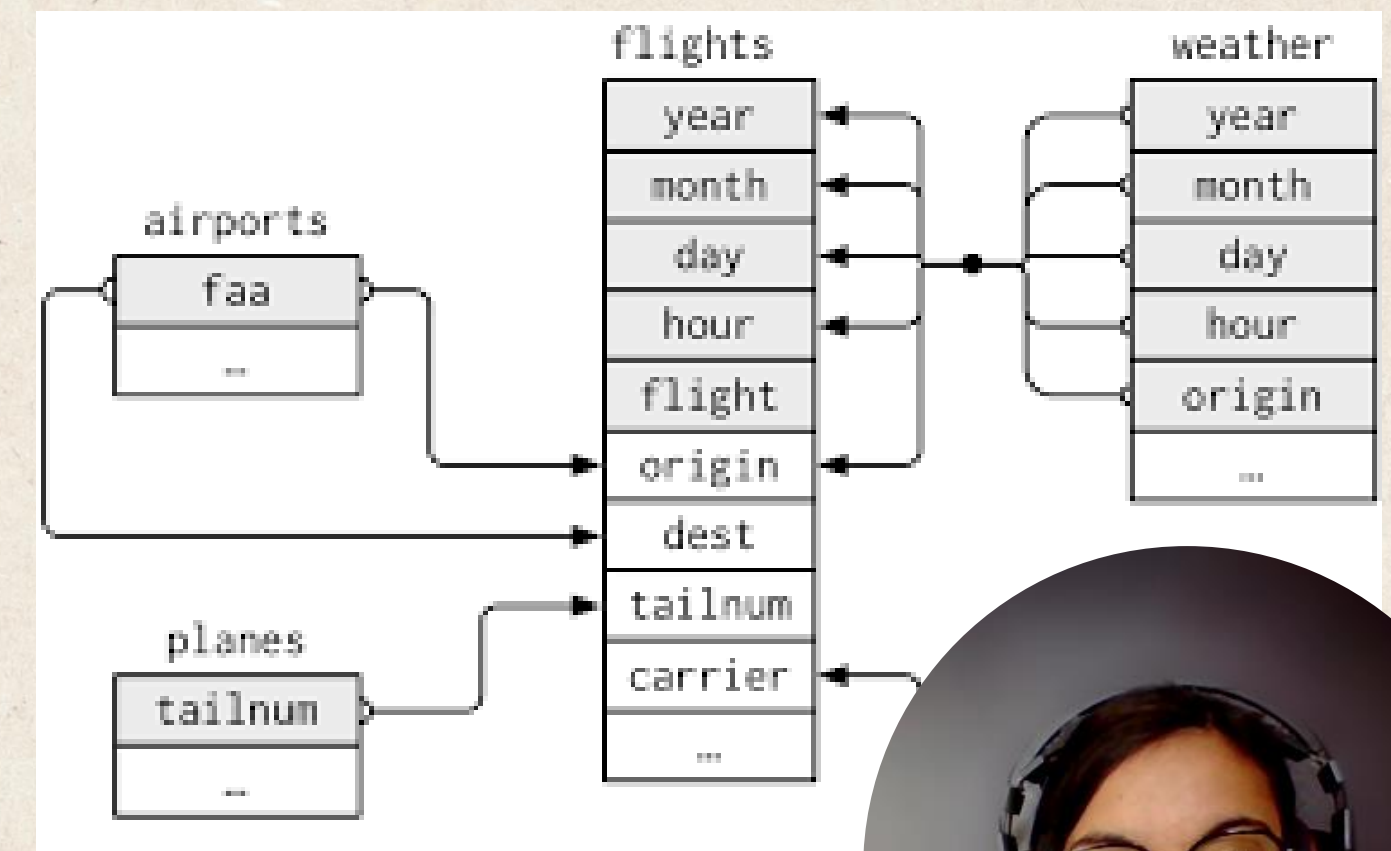
GitHub: <https://github.com/AndreaLaLupe/DataScienceProject/tree/main>

YouTube: <https://youtu.be/e87Ueufw-ag>

Descripción:

nycflights13 es un dataset que contiene información sobre vuelos que salieron de los aeropuertos de la ciudad de Nueva York en el año 2013.

- flights: contiene información detallada sobre vuelos individuales.
- airports: proporciona información sobre aeropuertos.
- airlines: contiene información sobre aerolíneas.
- planes: proporciona detalles sobre aviones individuales,
- weather: contiene datos relacionados con el clima para cada aeropuerto en cada día del año 2013



Sistema transaccional:

aws

Servicios

Buscar

[Alt+S]

Amazon RDS

Panel

Bases de datos

Editor de consultas

Información sobre rendimiento

Instantáneas de

Exportaciones en Amazon S3

Copias de seguridad automatizadas

Instancias reservadas

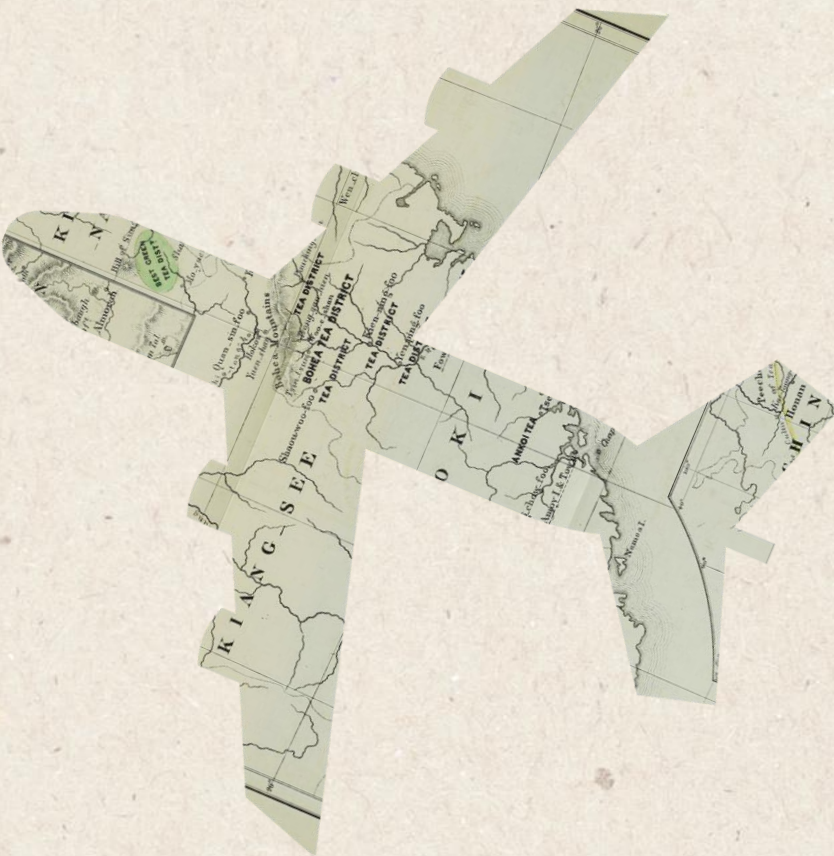
RDS > Bases de datos > nycflights13-database

nycflights13-database

Modificar Acciones

Resumen

Identificador de base de datos nycflights13-database	Estado Disponible	Rol Instancia	Motor MySQL Community	Recomendaciones 4 Informativo
CPU 3.41%	Clase db.t3.micro	Actividad actual 0	Región y AZ us-east-2b	
		Conexiones		



aws

Servicios

Buscar

[Alt+S]

Reglas de entrada

Información

ID de la regla del grupo de seguridad	Tipo	Protocolo	Intervalo de puertos	Origen	Descripción: opcional
	Información	Información	Información	Información	Información
sgr-0da04b98c01857650	Todo el tráfico	Todo	Todo	Per...	
sgr-076519ad03cd50fe0	MySQL/Aurora	TCP	3306	Per...	

Agregar regla



Sistema transaccional:

Connection "nycflights13_db" configuration

MySQL connection settings

Connection settings
Initialization
Shell Commands
Client identification
Transactions
General
Metadata
Errors and timeouts
Data Transfer
Data Editor
SQL Editor

Main | Driver properties | SSH | SSL

+ Network configurations...

Server
Connect by: ☒ Host ☐ URL
URL: jdbc:mysql://nycflights13-database.cvqmkq66ujho.us-east-2.rds.amazonaws.
Server Host: nycflights13-database.cvqmkq66ujho.us-east-2.rds.amazonaws Port: 3306
Database: nycflights13_db
Authentication (Database Native)
Username: admin
Password: ☒ Save password
Advanced
Server Time Zone: Auto-detect
Local Client: MySQL Binaries
[You can use variables in connection parameters.](#)
Driver name: MySQL
Driver Settings Driver license
Test Connection ... OK Cancel

Connection "nycflights13_db" configuration

MySQL connection settings

Connection settings
Initialization
Shell Commands
Client identification
Transactions
General
Metadata
Errors and timeouts
Data Transfer
Data Editor
SQL Editor

Main | Driver properties | SSH | SSL

+ Network configurations...

Server
Connect by: ☒ Host ☐ URL
URL: jdbc:mysql://nycflights13-database.cvqmkq66ujho.us-east-2.rds.amazonaws.
Server Host: nycflights13-database.cvqmkq66ujho.us-east-2.rds.amazonaws Port: 3306
Database: nycflights13_db
Authentication (Database Native)
Username: admin
Password: ☒ Save password
Advanced
Server Time Zone: Auto-detect
Local Client: MySQL Binaries
[You can use variables in connection parameters.](#)
Driver name: MySQL
Driver Settings Driver license
Test Connection ... OK Cancel

Connection test
Connected (1656 ms)
Server: MySQL 8.0.35
Driver: MySQL Connector/J mysql-connector-j-8.2.0 (Revision: 06a1f724497fd81c6a659131fda822c9e5085b6c)
OK Details >>
Server Time Zone: Auto-detect
Local Client: MySQL Binaries
[You can use variables in connection parameters.](#)
Driver name: MySQL
Driver Settings Driver license



DBeaver 24.0.0 - <nycflights13_db> nycflights_db.sql

File Edit Navigate Search SQL Editor Database Window Help

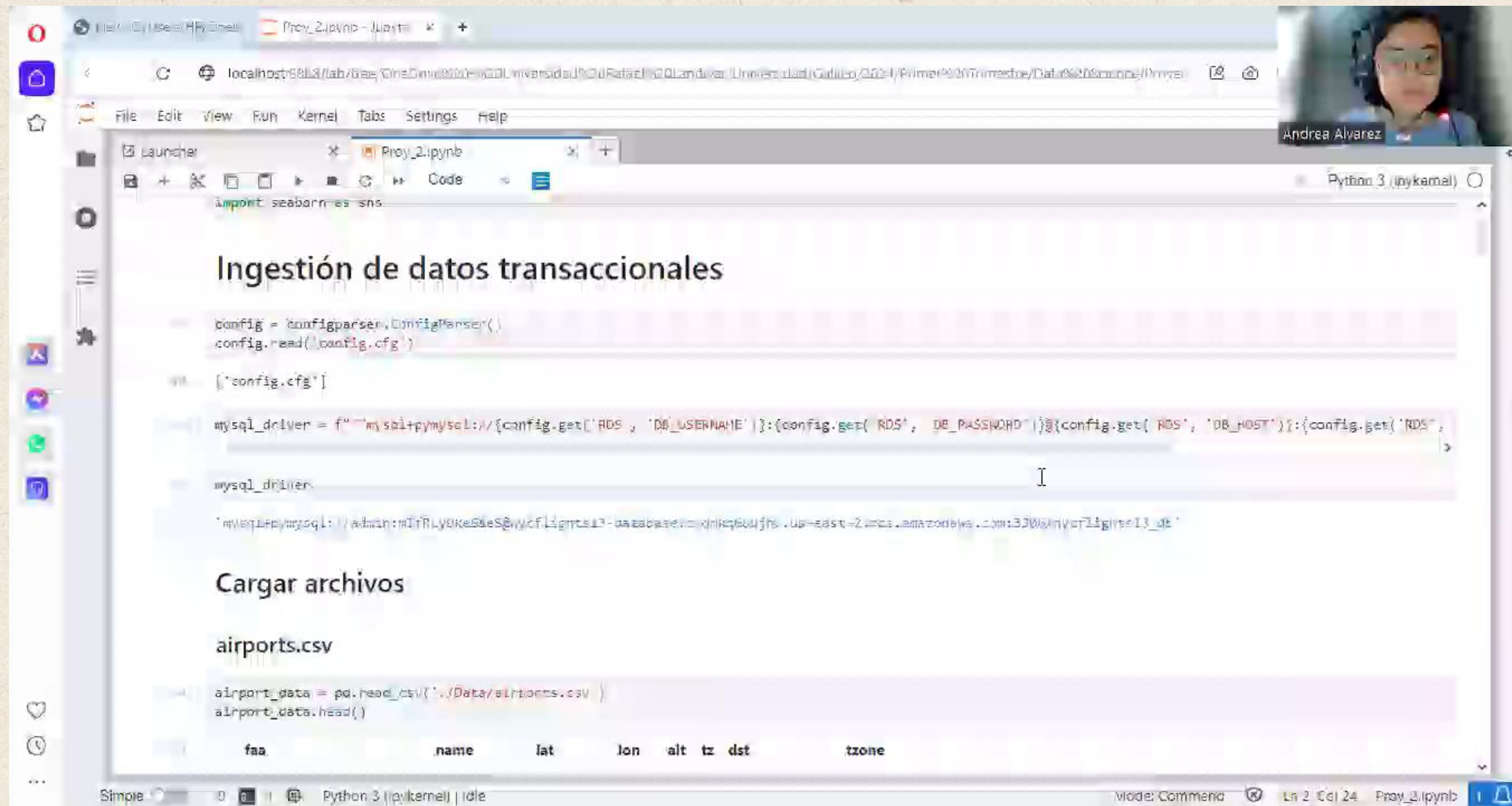
Database Navigator
Enter a part of object name here
nycflights13_db - nycflights13-database.cvqmkq66ujho.us-east-2.rds
Databases
nycflights13_db
Tables
airlines 16K
airports 160K
flights 97M
planes 352K
weather 3.5M
Views
Indexes
Procedures
Triggers
Events
sys
Users
Administer
System Info
sakila_db - db-rds-mysql.cvqmkq66ujho.us-east-2.rds.amazonaws.com

SQL Editor
nycflights13_db
nycflights13_db
Auto
Commit
Rollback
SQL
CREATE TABLE `airports` (
 `faa` varchar(3) NOT NULL,
 `name` varchar(100) NOT NULL,
 `lat` double NOT NULL,
 `lon` double NOT NULL,
 `alt` int NOT NULL,
 `tz` varchar(50) NOT NULL,
 `dst` varchar(1) NOT NULL,
 `tzone` varchar(100) NOT NULL,
 PRIMARY KEY (`faa`))
Statistics 1
Name Value
Updated Rows 0
Query
CREATE TABLE `weather` (
 `origin` varchar(3) NOT NULL,
 `year` int NOT NULL,
 `month` int NOT NULL,
 `day` int NOT NULL,
 `hour` int NOT NULL,
 `temp` double NOT NULL,
 PRIMARY KEY (`origin`, `year`, `month`, `day`, `hour`))

Output
Enter a part of a message to search



Ingestión de datos transaccionales:



```
import seaborn as sns

Ingestión de datos transaccionales

config = configparser.ConfigParser()
config.read('config.cfg')

['config.cfg']

mysql_driver = f"mysql+pymysql://{config.get('RDS', 'DB_USERNAME')}:{config.get('RDS', 'DB_PASSWORD')}@{config.get('RDS', 'DB_HOST')}:{config.get('RDS', 'DB_PORT')}"

mysql_driver

'mysql+pymysql://admin:W17R1Y0KES8eSgWjF1g1rte13-database:cn1qy6u4jfr.us-east-2.amazonaws.com:3306/W17R1Y0KES8eSgWjF1g1rte13_db'

Cargar archivos

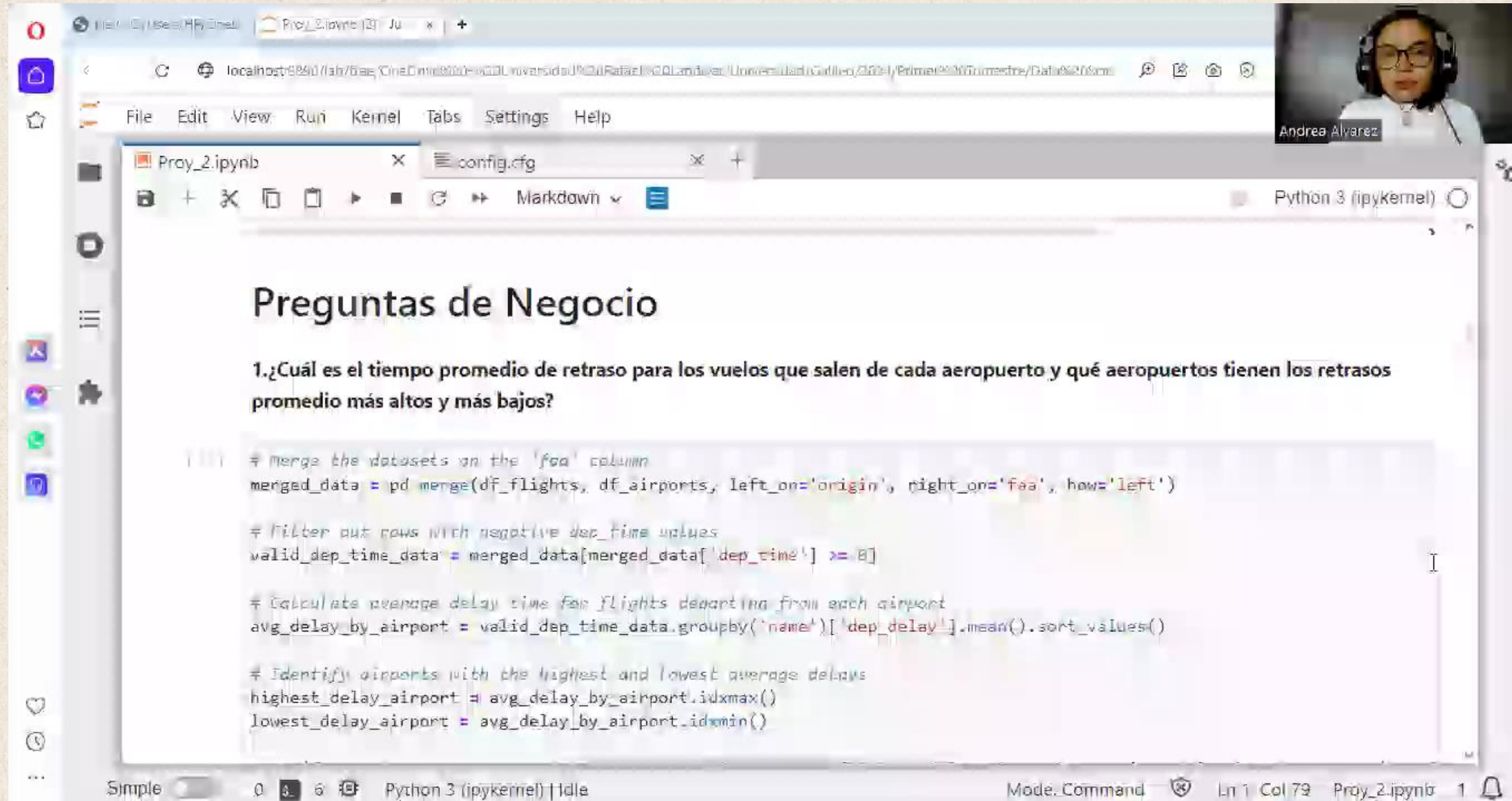
airports.csv

airport_data = pd.read_csv('../Data/airports.csv')
airport_data.head()
```

faa	name	lat	lon	alt	tz	dst	tzone
-----	------	-----	-----	-----	----	-----	-------



Preguntas de Negocio:



The screenshot shows a Jupyter Notebook interface with a video call window in the top right corner. The video call window shows a person named Andrea Alvarez. The Jupyter Notebook has a tab titled "Proy_2.ipynb" and a "config.cfg" file. The notebook content includes the title "Preguntas de Negocio" and a list of business questions. The first question is: "1. ¿Cuál es el tiempo promedio de retraso para los vuelos que salen de cada aeropuerto y qué aeropuertos tienen los retrasos promedio más altos y más bajos?". Below the question is a code cell containing Python code that merges two datasets, filters out rows with negative departure times, calculates the average delay time for flights departing from each airport, and identifies the airports with the highest and lowest average delays.

```
# Merge the datasets on the 'faa' column
merged_data = pd.merge(df_flights, df_airports, left_on='origin', right_on='faa', how='left')

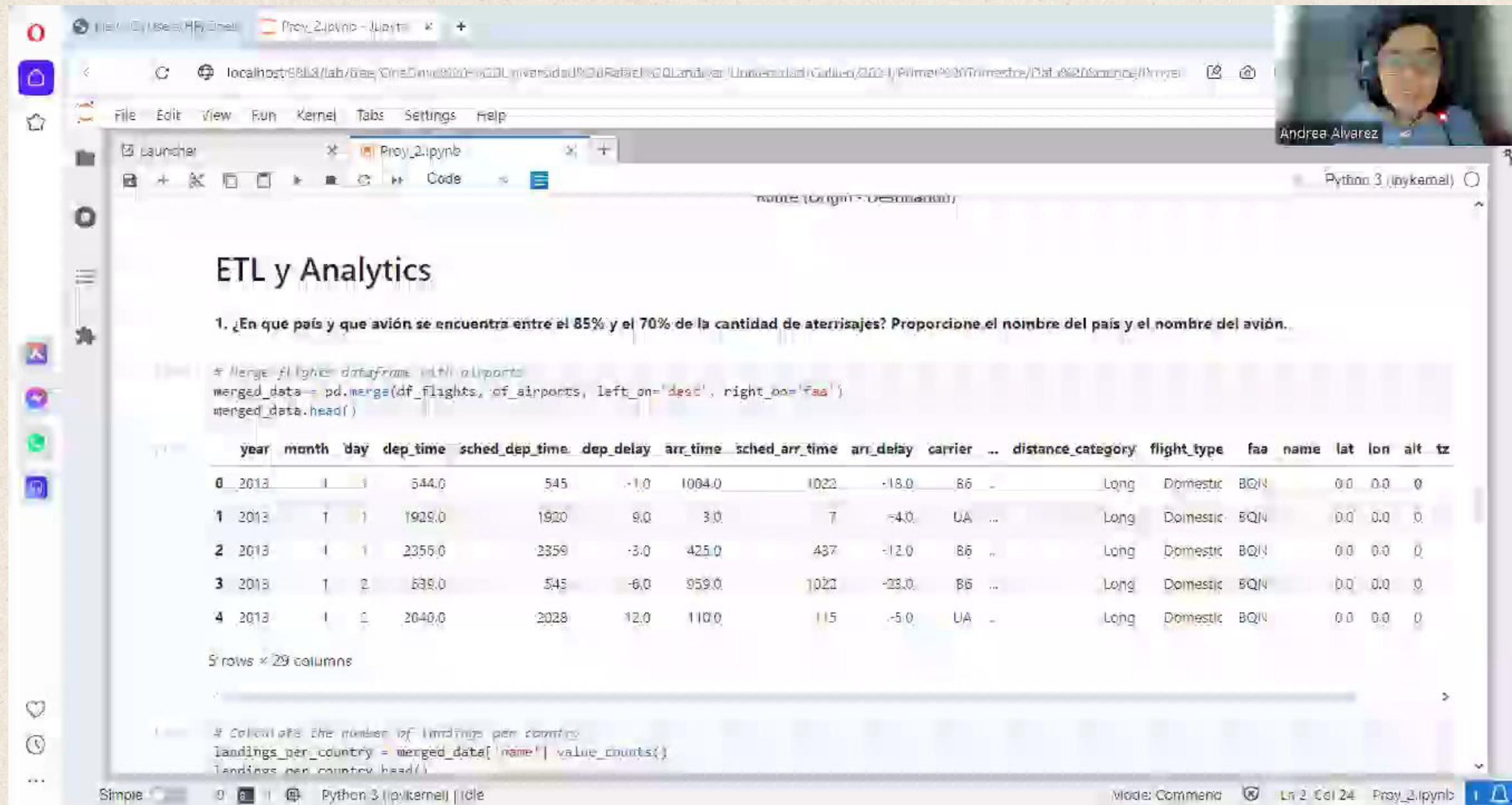
# Filter out rows with negative dep_time values
valid_dep_time_data = merged_data[merged_data['dep_time'] >= 0]

# Calculate average delay time for flights departing from each airport
avg_delay_by_airport = valid_dep_time_data.groupby('name')['dep_delay'].mean().sort_values()

# Identify airports with the highest and lowest average delays
highest_delay_airport = avg_delay_by_airport.idxmax()
lowest_delay_airport = avg_delay_by_airport.idxmin()
```



ETL y Analytics:



ETL y Analytics

1. ¿En qué país y que avión se encuentra entre el 85% y el 70% de la cantidad de aterrisajes? Proporcione el nombre del país y el nombre del avión.

```
# Merge flight dataframe with airports
merged_data = pd.merge(df_flights, df_airports, left_on='dest', right_on='faa')
merged_data.head()
```

	year	month	day	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time	arr_delay	carrier	...	distance_category	flight_type	faa	name	lat	lon	alt	tz
0	2013	1	1	544.0	545	-1.0	1004.0	1022	-18.0	B6	...	Long	Domestic	BQN	0.0	0.0	0		
1	2013	1	1	1929.0	1920	9.0	3.0	7	-4.0	UA	...	Long	Domestic	BQN	0.0	0.0	0		
2	2013	1	1	2356.0	2359	-3.0	425.0	437	-12.0	B6	...	Long	Domestic	BQN	0.0	0.0	0		
3	2013	1	2	539.0	545	-6.0	959.0	1022	-23.0	B6	...	Long	Domestic	BQN	0.0	0.0	0		
4	2013	1	2	2040.0	2028	12.0	110.0	115	-5.0	UA	...	Long	Domestic	BQN	0.0	0.0	0		

5 rows x 29 columns

```
# Calculate the number of landings per country
landings_per_country = merged_data['name'].value_counts()
landings_per_country.head()
```

Mode: Command | Ln 2 | Col 24 | Proj_3.ipynb



GitHub:

Git

```
: %%bash
git init
git add .
git commit -m "Upload Proyecto2aParte"
git remote add origin https://github.com/AndreaLaLupe/DataScienceProject.git
git push -u origin main
```

AndreaLaLupe / DataScienceProject

Type to search

>

+

🔍

🔗

📧

🏠

<> Code

🕒 Issues

🔗 Pull requests

🎮 Actions

📁 Projects

📖 Wiki

🛡 Security

📈 Insights

⚙ Settings

🏠 DataScienceProject

Public

📌 Pin

👁 Unwatch 1

🍴 Fork 0

★ Star 0

🔗 main

🔗 1 Branch

🏷 0 Tags

🔍 Go to file

⌵ Add file

<> Code

About

👤 Your Name Upload Proyecto2aParte 1d4037b · now 11 Commits

📁 .ipynb_checkpoints

Upload Proyecto Segunda Parte

1 hour ago

📁 Data

Upload Proyecto Segunda Parte

1 hour ago

📄 Data Science Project.pptx

Upload Proyecto2aParte

now

📄 Proy_2.ipynb

Upload Proyecto2aParte

now

📄 Proyecto_Final.pdf

Upload Proyecto Segunda Parte

1 hour ago

📄 config.cfg

Upload Proyecto Segunda Parte

1 hour ago

📄 data.zip

Upload Proyecto Segunda Parte

1 hour ago

📄 nycflights_db.sql

Upload Proyecto2aParte

5 minutes ago

📄 Activity

★ 0 stars

👁 1 watching

🍴 0 forks

Releases

No releases published

Create a new release

Packages

No packages published

Publish your first package





Muchas gracias!

Andrea Alvarez - 23003779

