

nycflights13

Data Science

Link

<https://youtu.be/XogtmTsVAdw>

A collage background featuring torn map pieces, a paper airplane, and architectural drawings. The maps show various geographical features and place names. The paper airplane is green and white. The architectural drawings include a detailed drawing of a building facade and a map of Japan labeled 'IAPON'.

Link

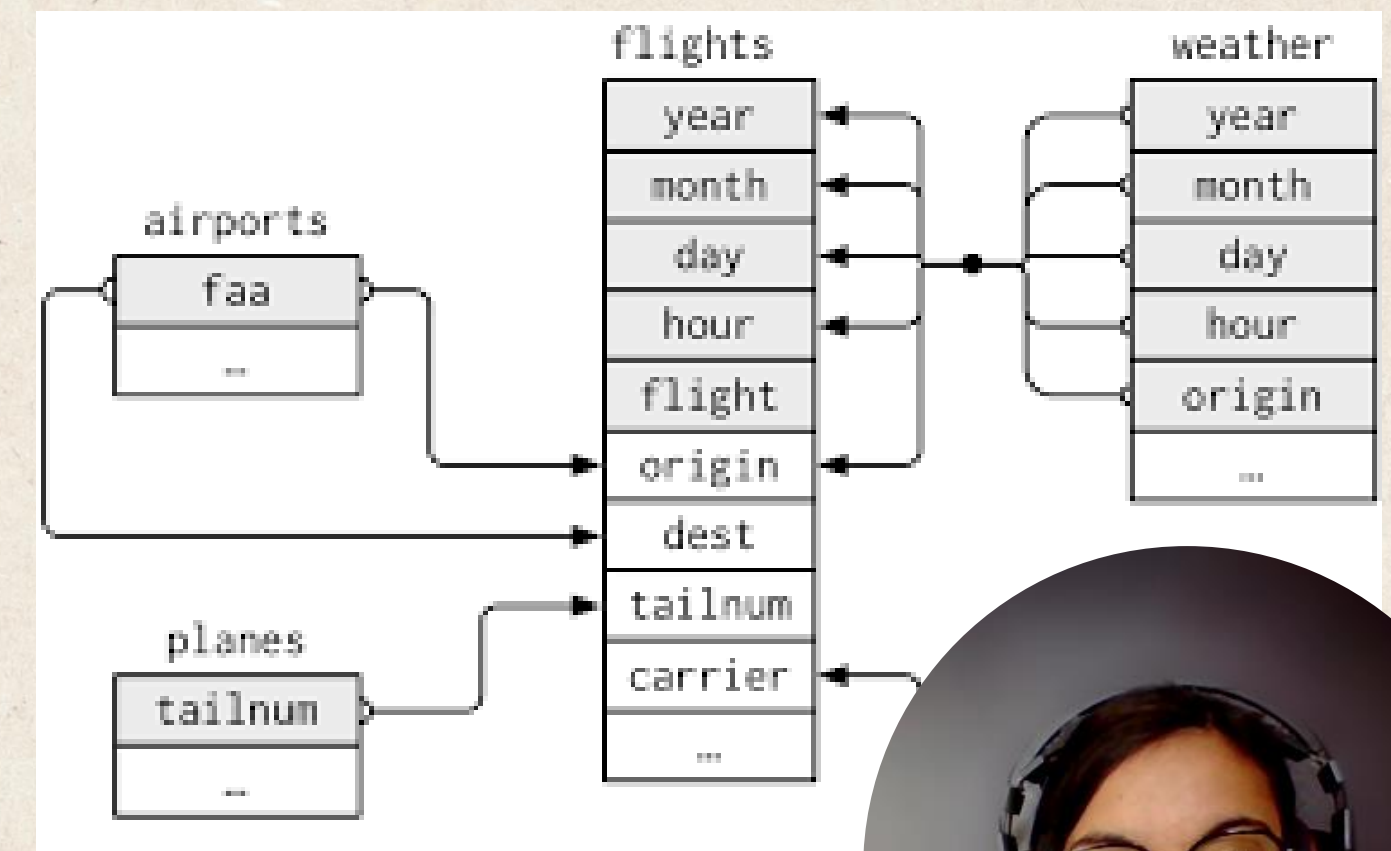
<https://youtu.be/XogtmTsVAdw>



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

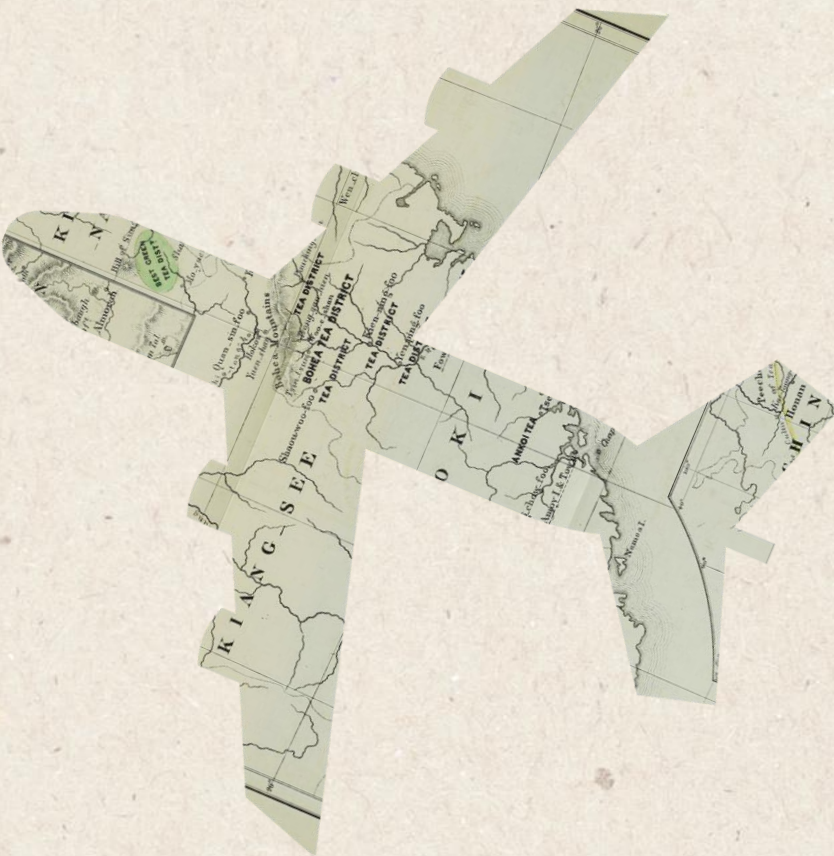
RDS > Bases de datos > nycflights13-database

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 >>

DBeaver 24.0.0 - <nycflights13_db> nycflights_db.sql

File Edit Navigate Search SQL Editor Database Window Help

Database Navigator

Projects

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

<nycflights13_db> nycflights_db.sql

DROP TABLE IF EXISTS `airports`;

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,

DDTMADV 22V ('faa')

Statistics 1

Name

Value

Updated Rows

0

Query

CREATE TABLE `weather` (

`origin` varchar(3) NOT NULL,

`year` int NOT NULL,

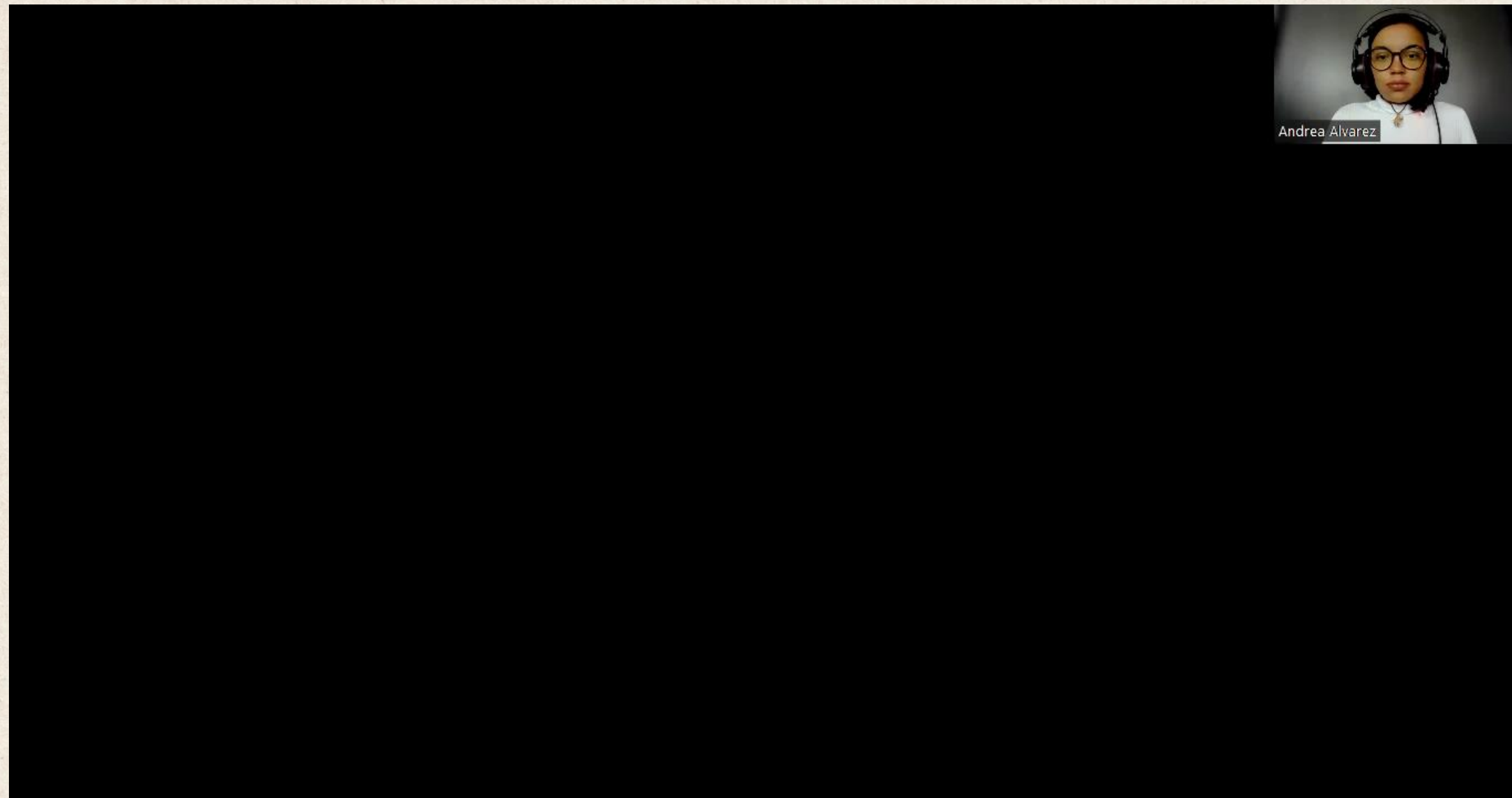
`month` int NOT NULL,

`day` int NOT NULL,

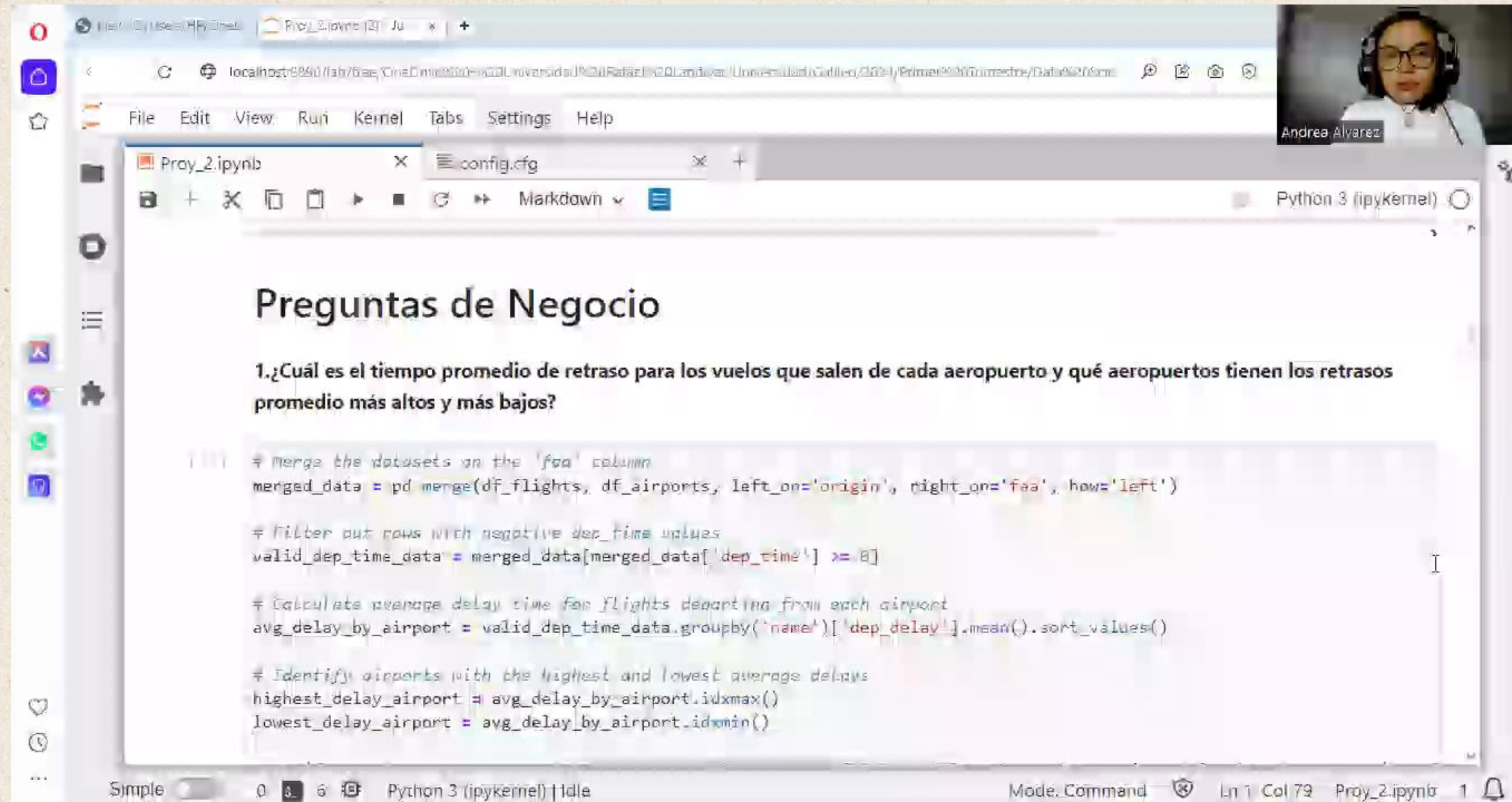
`hour` int NOT NULL,

`temp` double NOT NULL,

Ingestión de datos transaccionales:



Preguntas de Negocio:



The screenshot shows a Jupyter Notebook window titled 'Proy_2.ipynb'. The browser address bar shows a local file path. The notebook interface includes a menu bar (File, Edit, View, Run, Kernel, Tabs, Settings, Help) and a toolbar with icons for file operations and execution. The main content area displays the title 'Preguntas de Negocio' followed by a business question and a code cell.

Preguntas de Negocio

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?

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

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

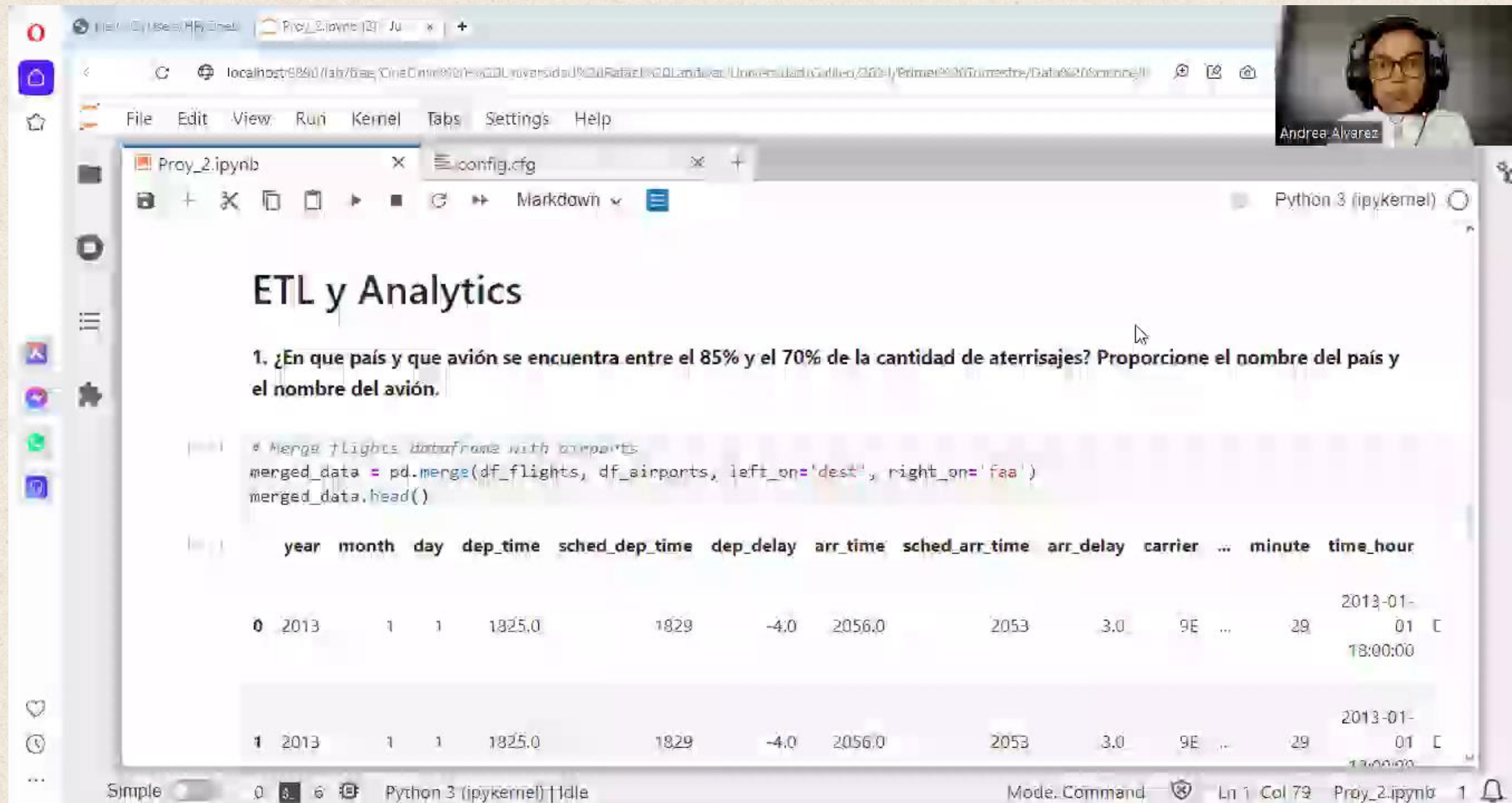
# Calculate average delay time for flights departing from each airport
avg_delay_by_airport = valid_dep_time_data.groupby('name')['dep_delay'].mean().sort_values()

# 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()
```

The bottom status bar indicates 'Simple' mode, 'Python 3 (ipykernel)' as the kernel, and 'Mode: Command' with cursor coordinates 'Ln 1 Col 79'.



ETL y Analytics:



The screenshot shows a Jupyter Notebook interface with a browser window at the top displaying the URL: `localhost:8840/lab?base=0&CineOnline%20Herramientas%20de%20Análisis%20de%20Datos%20de%20Vuelos%20de%20América%20Latina%20y%20el%20Caribe%202013%20I%20Trimestre/Data%20Science/`. The notebook has two tabs: `Proy_2.ipynb` and `config.cfg`. The main content area shows the title **ETL y Analytics** and a task: **1. ¿En que 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.**

Below the task, there is a code cell with the following Python code:

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

The output of the code is a table with the following columns: `year`, `month`, `day`, `dep_time`, `sched_dep_time`, `dep_delay`, `arr_time`, `sched_arr_time`, `arr_delay`, `carrier`, `...`, `minute`, and `time_hour`. The table shows two rows of data:

	year	month	day	dep_time	sched_dep_time	dep_delay	arr_time	sched_arr_time	arr_delay	carrier	...	minute	time_hour
0	2013	1	1	1825.0	1829	-4.0	2056.0	2053	3.0	9E	...	29	2013-01-18:00:00
1	2013	1	1	1825.0	1829	-4.0	2056.0	2053	3.0	9E	...	29	2013-01-18:00:00

The bottom status bar indicates the notebook is running on `Python 3 (ipykernel)` and is in `Mode: Command`. The cursor is at `Ln 1 Col 79` in the `Proy_2.ipynb` file.



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

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

About

Project of the course Data Science

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

