

Billones, Cristel Kaye  
Rosales, Frances Aneth  
Samson, Justine Aizel D.

FA 4

## Big Data Analytics

### Screenshots of the Export Results

The screenshot shows a Google Colab notebook titled "FA4\_Billones\_Rosales\_Samson\_BDA.ipynb". The left sidebar displays the "Table of contents" with sections: "FA 4", "Importing Libraries", "Hands-on Coding Task", "Importing Dataset", "Cleaning Dataset and Displaying First 10 Rows", "Transform and analyze the data", "Export Results", and "+ Section". The main area shows a table of data with columns for date, distance, and duration. Below the table, "Key Observations" are listed: highest total distance on 2025-01-16, longest average trip on 2025-01-02, shortest average trip on 2025-01-19, and low activity day on 2024-12-31. The "Export Results" section contains two code blocks: one for exporting to Parquet and another for exporting to CSV.

Date	Distance (miles)	Duration (minutes)
2025-01-17	940059.8699000076	10.006750181707808
2025-01-18	474121.8100000079	14.5821872557618
2025-01-19	617234.7200000090	15.02799918000020

only showing top 30 rows

**Key Observations:**

- Highest total distance: 2025-01-16 with 1,062,429.64 miles, suggesting peak taxi activity on that day.
- Longest average trip: 2025-01-02 with 16.92 minutes.
- Shortest average trip: 2025-01-19 with 13.02 minutes, despite a relatively high total distance (~637k miles), indicating many short trips.
- Low activity day: 2024-12-31 only had 76.01 miles — likely a partial day (perhaps late data start).

**Export Results**

```
[12] output_path = "/content/drive/My Drive/Colab Notebooks/yellow_tripdata_2025-01_cleaned"
df_clean.write_mode("overwrite").parquet(output_path)

[13] csv_output_path = "/content/drive/My Drive/Colab Notebooks/yellow_tripdata_2025-01_cleaned_csv"
df_clean.write_mode("overwrite").option("header", True).csv(csv_output_path)
```

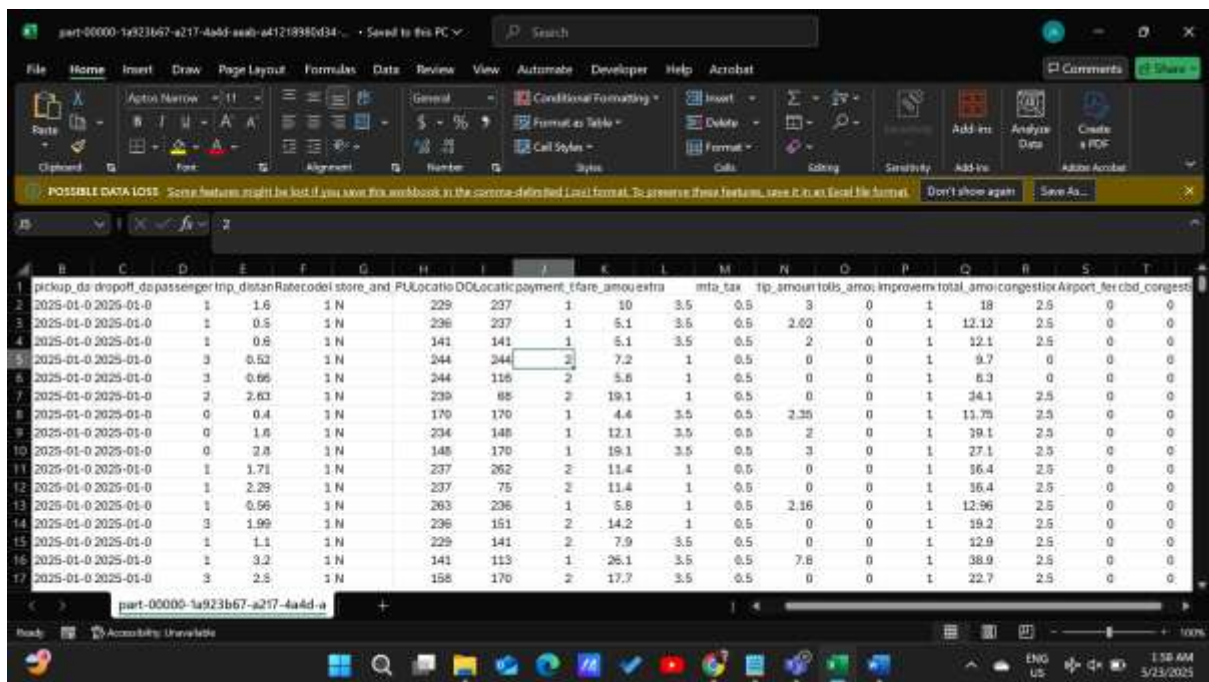
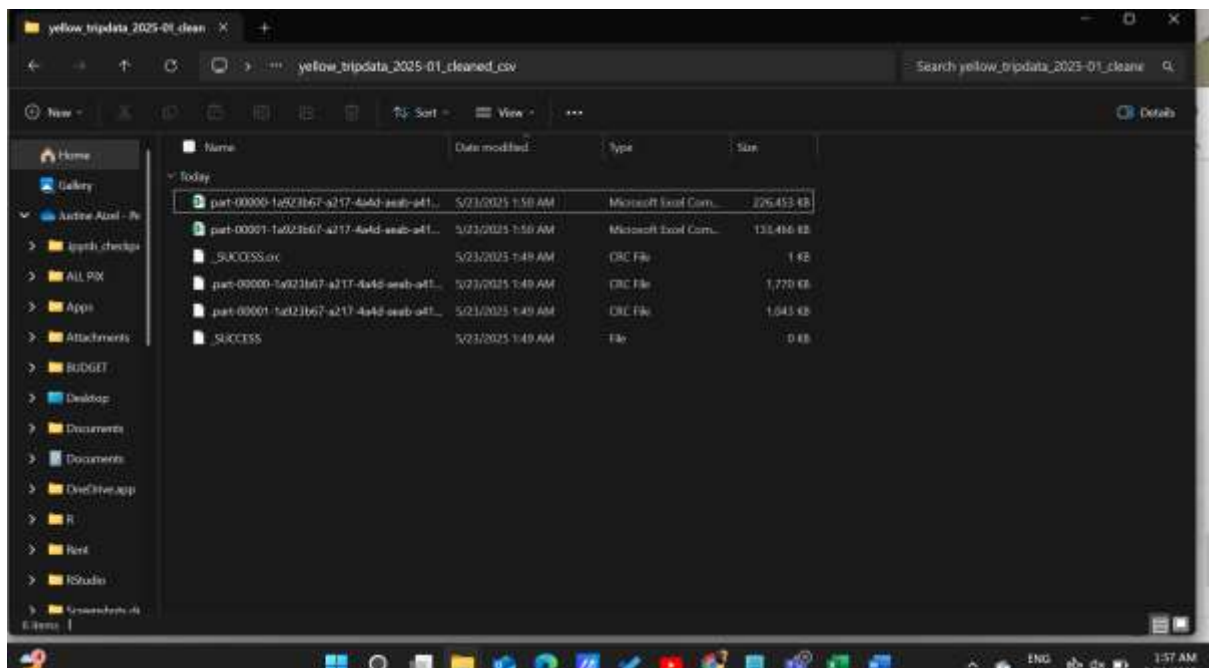
The screenshot shows the Google Drive interface. The left sidebar shows the navigation pane with "My Drive" selected. The main area shows the "Colab Notebooks" folder. Three files are listed: "yellow\_tripdata\_2025-01\_csv", "yellow\_tripdata\_2025-01\_cleaned\_csv", and "yellow\_tripdata\_2025-01\_cleaned".

Name	Owner	Last modified	File size
yellow_tripdata_2025-01_csv	me	1:06 AM	—
yellow_tripdata_2025-01_cleaned_csv	me	1:39 AM	—
yellow_tripdata_2025-01_cleaned	me	1:38 AM	—

### CSV File Type

The screenshot shows a Windows File Explorer window. The address bar shows the path: "Downloads > yellow\_tripdata\_2025-01\_cleaned\_csv-20250522T174720Z-1-001". The left sidebar shows the navigation pane with "Documents" and "OneDrive.app" selected. The main area shows a list of files and folders. A folder named "yellow\_tripdata\_2025-01\_cleaned\_csv" is highlighted.

Name	Date modified	Type	Size
yellow_tripdata_2025-01_cleaned_csv	5/23/2025 1:50 AM	File folder	



part-00001-1a923b67-a217-4a4d-aabb-a41219980b34... - Saved to this PC

Search

File Home Insert Draw Page Layout Formulas Data Review View Automate Developer Help Acrobat

Clipboard Font Paragraph Alignment Number Styles Conditional Formatting Insert Delete Format Cells Editing Sensitivity Add-ins Analyze Data Create a PDF Adobe Acrobat

POSSIBLE DATA LOSS: Some features might be lost if you save this workbook in the compacted binary format. To preserve these features, save it as an Excel file format. Don't show again Save As...

A1 VendorID

VendorID	pickup_datetime	dropoff_datetime	passenger_count	trip_distance	rate_code	store_and_fwd_flag	PULocationID	DOLocationID	payment_type	fare_amount	extra	mta_tax	tip_amount	tolls_amount	improvement_surcharge	total_amount	congestion_surcharge	airport_fee	city_state_country
2	2025-01-2 2025-01-2	1	2.85	1	N		230	239	1	15.5	2.5	0.5	5.71	0	1	28.56	2.5	0	
2	2025-01-2 2025-01-2	1	0.48	1	N		162	162	1	5.8	2.5	0.5	1.5	0	1	14.55	2.5	0	
2	2025-01-2 2025-01-2	1	3.16	1	N		162	79	1	16.3	2.5	0.5	4.71	0	1	28.26	2.5	0	
2	2025-01-2 2025-01-2	1	1.37	1	N		234	144	1	12.8	2.5	0.5	4.01	0	1	24.66	2.5	0	
2	2025-01-2 2025-01-2	1	2.06	1	N		68	107	1	13.5	2.5	0.5	4.15	0	1	24.9	2.5	0	
2	2025-01-2 2025-01-2	2	1.45	1	N		107	4	1	12.1	2.5	0.5	3.87	0	1	23.22	2.5	0	
2	2025-01-2 2025-01-2	2	21.07	2	N		132	261	2	70	5	0.5	0	8.94	1	88.44	2.5	1.75	
2	2025-01-2 2025-01-2	6	9.48	1	N		132	39	1	40.1	2.5	0.5	5	0	1	50.85	0	1.75	
1	2025-01-2 2025-01-2	1	0.2	1	N		107	137	1	3.7	5.75	0.5	2.75	0	1	13.7	2.5	0	
2	2025-01-2 2025-01-2	1	1.88	1	N		234	137	1	14.9	2.5	0.5	4.43	0	1	26.58	2.5	0	
2	2025-01-2 2025-01-2	1	0.61	1	N		137	106	1	13.5	2.5	0.5	4.15	0	1	24.9	2.5	0	
2	2025-01-2 2025-01-2	1	0.61	1	N		164	233	1	10	2.5	0.5	1	0	1	18.25	2.5	0	
2	2025-01-2 2025-01-2	2	1.42	1	N		233	230	1	12.1	2.5	0.5	3.87	0	1	23.22	2.5	0	
2	2025-01-2 2025-01-2	2	1.32	1	N		231	249	1	11.4	2.5	0.5	3.73	0	1	22.38	2.5	0	
2	2025-01-2 2025-01-2	2	3.37	1	N		158	229	1	26.8	2.5	0.5	5	0	1	39.05	2.5	0	
2	2025-01-2 2025-01-2	1	1.82	1	N		161	107	1	14.9	2.5	0.5	5.54	0	1	27.69	2.5	0	

part-00001-1a923b67-a217-4a4d-a

Accessibility Unavailable

ENG US 1:55 AM 5/23/2025

## Parquet File Type

yellow\_tripdata\_2025-01\_clean

Search yellow\_tripdata\_2025-01\_clean

File View Sort View Details

Name	Date modified	Type	Size
part-00000-54410a01-5a0c-47e5-8dae-f5c...	5/23/2025 2:05 AM	PARQUET file	41,812 KB
_SUCCESS.txt	5/23/2025 2:05 AM	ORC file	1 KB
part-00000-54410a01-5a0c-47e5-8dae-f5c...	5/23/2025 2:05 AM	ORC file	321 KB
part-00001-54410a01-5a0c-47e5-8dae-f5c...	5/23/2025 2:05 AM	ORC file	189 KB
_SUCCESS	5/23/2025 2:05 AM	File	0 KB
part-00001-54410a01-5a0c-47e5-8dae-f5c...	5/23/2025 2:05 AM	PARQUET file	24,161 KB

6 items

ENG US 2:07 AM 5/23/2025

Opened it through online parquet viewer

taxi.parquet?file-id=tbl\_1747937145541\_5965w3b

Select Aggregate SQL Run Query (Ctrl+Enter) Click a query type to expand Export

Nova AI

Schema Information

part\_00000\_54410e81\_5e0c\_43e5\_8dae\_f

Try one of these:

- Analyze my data trends  
Discover patterns and insights
- Find unusual patterns  
Detect outliers and anomalies
- Quick data summary  
Get key statistics instantly

Upgrade to use Nova AI

Upgrade to our Nova AI

	VendorID	pickup_datetime	dropoff_datetime	passenger_count	trip_distance	RatecodeID
1	1	1735690718000	1735691219000	1	1.6	1
1	1	1735691560000	1735691713000	1	0.5	1
1	1	1735692244000	1735692361000	1	0.6	1
2	2	1735690467000	1735690801000	3	0.32	1
2	2	1735690894000	1735691106000	3	0.66	1
2	2	1735692504000	1735693706000	2	2.63	1
1	1	1735690487000	1735690575000	0	0.4	1
1	1	1735691967000	1735692711000	0	1.6	1
1	1	1735692823000	1735694003000	0	2.8	1
2	2	1735689602000	1735690176000	1	1.71	1
2	2	1735690828000	1735691284000	1	2.29	1
2	2	1735691638000	1735691843000	1	0.56	1
2	2	1735692160000	1735692938000	3	1.99	1
1	1	1735691407000	1735691808000	1	1.1	1

by @ecottipaulin

1 2 ... 40