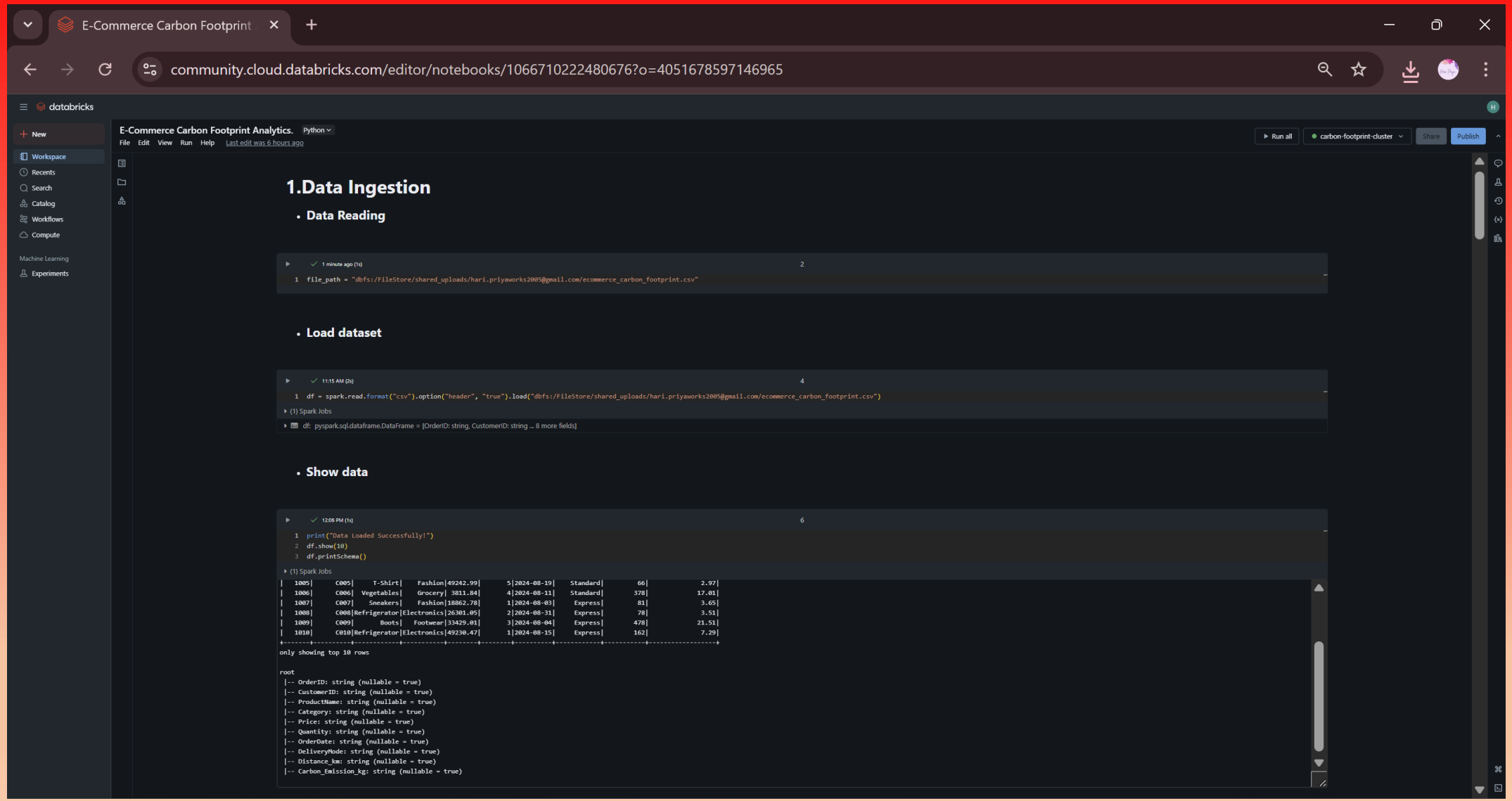
https://github.com/Haripriya2-2k5/E-Commerce-Carbon-Footprint-Analytics


E-Commerce Carbon Footprint
community.cloud.databricks.com/editor/notebooks/1066710222480676?o=4051678597146965
databricks
New
Workspace
Recents
Search
Catalog
Workflows
Compute
Machine Learning
Experiments
E-Commerce Carbon Footprint Analytics.
Python
File Edit View Run Help
Last edit was 6 hours ago
Run all
carbon-footprint-cluster
Share
Publish
1.Data Ingestion
Data Reading
1 minute ago (1s)
2
1 file_path = "dbfs:/FileStore/shared_uploads/hari.priyaworks2005@gmail.com/ecommerce_carbon_footprint.csv"
Load dataset
11:15 AM (2s)
4
1 df = spark.read.format("csv").option("header", "true").load("dbfs:/FileStore/shared_uploads/hari.priyaworks2005@gmail.com/ecommerce_carbon_footprint.csv")
(1) Spark Jobs
df: pyspark.sql.dataframe.DataFrame = [OrderID: string, CustomerID: string ... 8 more fields]
Show data
12:06 PM (1s)
6
1 print("Data Loaded Successfully!")
2 df.show(10)
3 df.printSchema()
(1) Spark Jobs
1005	C005	T-Shirt	Fashion	49242.99	5	2024-08-19	Standard	66	2.97
1006	C006	Vegetables	Grocery	3811.84	4	2024-08-11	Standard	378	17.01
1007	C007	Sneakers	Fashion	18862.78	1	2024-08-03	Express	81	3.65
1008	C008	Refrigerator	Electronics	26301.05	2	2024-08-31	Express	78	3.51
1009	C009	Boots	Footwear	33429.01	3	2024-08-04	Express	478	21.51
1010	C010	Refrigerator	Electronics	49230.47	1	2024-08-15	Express	162	7.29
only showing top 10 rows
root
-- OrderID: string (nullable = true)
-- CustomerID: string (nullable = true)
-- ProductName: string (nullable = true)
-- Category: string (nullable = true)
-- Price: string (nullable = true)
-- Quantity: string (nullable = true)
-- OrderDate: string (nullable = true)
-- DeliveryMode: string (nullable = true)
-- Distance_km: string (nullable = true)
-- Carbon_Emission_kg: string (nullable = true)

E-Commerce Carbon Footprint

community.cloud.databricks.com/editor/notebooks/1066710222480676?o=4051678597146965

2.Data Cleaning

- **Convert Spark DataFrame to Pandas for easy visualization**

```
11:21 AM (1s)                                                    9
1  pdf = df.toPandas()
(1) Spark Jobs
```

- **Check for missing values**

```
11:23 AM (<1s)                                                   11
1  print("Missing values:\n", pdf.isnull().sum())
```

```
Missing values:
 OrderID          0
CustomerID        0
ProductName       0
Category          0
Price             0
Quantity          0
OrderDate         0
DeliveryMode      0
Distance_km       0
Carbon_Emission_kg    0
dtype: int64
```

- **Drop rows with nulls if any**

```
11:23 AM (<1s)                                                   13
1  pdf.dropna(inplace=True)
```

- **Verify cleaning**

```
11:27 AM (<1s)                                                   15
1  print("Data after cleaning:", pdf.shape)
2  pdf.head()
```

```
Data after cleaning: (100, 10)
```

	OrderID	CustomerID	ProductName	Category	Price	Quantity	OrderDate	DeliveryMode	Distance_km	Carbon_Emission_kg
0	1001	C001	Curtains	Home	13175.78	3	2024-08-28	Express	94	4.23
1	1002	C002	Bread	Grocery	48245.4	2	2024-08-07	Standard	395	17.77
2	1003	C003	Sneakers	Footwear	4413.29	2	2024-08-25	Express	376	16.92
3	1004	C004	Sandals	Footwear	5059.11	5	2024-08-07	Standard	269	12.11
4	1005	C005	T-Shirt	Fashion	49242.99	5	2024-08-19	Standard	66	2.97

+ Code + Text

E-Commerce Carbon Footprint

community.cloud.databricks.com/editor/notebooks/1066710222480676?o=4051678597146965

databricks

New

Workspace

Recents

Search

Catalog

Workflows

Compute

Machine Learning

Experiments

E-Commerce Carbon Footprint Analytics. Python

File Edit View Run Help Last edit was 6 hours ago

Run all | carbon-footprint-cluster | Share | Publish

0	1001	C001	Curtains	Home	13175.78	3	2024-08-28	Express	94	4.23
1	1002	C002	Bread	Grocery	48245.4	2	2024-08-07	Standard	395	17.77
2	1003	C003	Sneakers	Footwear	4413.29	2	2024-08-25	Express	376	16.92
3	1004	C004	Sandals	Footwear	5059.11	5	2024-08-07	Standard	269	12.11
4	1005	C005	T-Shirt	Fashion	49242.99	5	2024-08-19	Standard	66	2.97

3.Carbon Footprint Calculation

- **Total carbon emissions**
- **Carbon Emission by Category**
- **Carbon Emission by Delivery Mode**

11:33 AM (<1s) 18

```
total_emission = pdf['Carbon_Emission_kg'].sum()
print(f"Total Carbon Emission: {total_emission} kg")
```

```
Total Carbon Emission: 4.2317.7716.9212.112.9717.013.653.5121.517.296.4320.756.4816.5618.591.574.7213.6313.3717.822.487.7410.570.856.176.086.081.35.8912.964.687.068.4117.468.289.040.4918.1822.234.6314.7111.613.3811.340.9416.5611.439.3116.0616.3819.4410.268.699.454.911.357.212.556.5718.367.24.9116.5214.9821.155.315.8910.2112.5512.7819.8911.470.916.5613.419.5414.260.9915.161.8920.722.06.0814.5415.6619.121.532.979.6710.815.484.0910.4822.52.8318.3618.223.335.811.43 kg
```

11:34 AM (<1s) 19

```
category_emission = pdf.groupby('Category')['Carbon_Emission_kg'].sum().sort_values(ascending=False)
print("\nCarbon Emission by Category:\n", category_emission)
```

```
Carbon Emission by Category:
 Category
Home           4.236.432.487.748.419.040.4918.1811.6116.0610....
Electronics    3.517.2916.560.856.081.312.9617.468.2822.2314....
Fashion        2.973.6520.751.574.7213.3717.826.174.6311.340....
Grocery        17.7717.0113.6310.576.085.894.687.063.389.4512...
Footwear       16.9212.1121.516.4818.598.697.26.5718.365.315....
Name: Carbon_Emission_kg, dtype: object
```

11:34 AM (<1s) 20

```
delivery_emission = pdf.groupby('DeliveryMode')['Carbon_Emission_kg'].sum()
print("\nCarbon Emission by Delivery Mode:\n", delivery_emission)
```

```
Carbon Emission by Delivery Mode:
 DeliveryMode
Express     4.2316.923.653.5121.517.296.4320.7516.5618.591...
Standard    17.7712.112.9717.016.484.7213.6313.372.487.746...
Name: Carbon_Emission_kg, dtype: object
```

E-Commerce Carbon Footprint

community.cloud.databricks.com/editor/notebooks/1066710222480676?o=4051678597146965

databricks

New

Workspace

Recents

Search

Catalog

Workflows

Compute

Machine Learning

Experiments

E-Commerce Carbon Footprint Analytics. Python

File Edit View Run Help Last edit was 6 hours ago

Run all carbon-footprint-cluster Share Publish

```
Standard    17.7712.112.9717.016.484.7213.6313.372.487.746...
Name: Carbon_Emission_kg, dtype: object
```

4.Visualization using Matplotlib and pandas

Bar chart – CO_2 emissions by delivery mode

11:55 AM (1s) 23

```
1
2  import matplotlib.pyplot as plt
3  import pandas as pd
4
5  # Convert Spark DataFrame to Pandas
6  pdf = df.toPandas()
7
8  # Group by DeliveryMode for visualization
9  mode_summary = pdf.groupby("DeliveryMode")["Carbon_Emission_kg"].sum().reset_index()
10
11 # Plot a bar chart
12 plt.figure(figsize=(8, 5))
13 plt.bar(mode_summary["DeliveryMode"], mode_summary["Carbon_Emission_kg"], color=['skyblue', 'lightgreen', 'orange'])
14 plt.title("CO₂ Emissions by Delivery Mode", fontsize=14)
15 plt.xlabel("Delivery Mode", fontsize=12)
16 plt.ylabel("Total CO₂ Emissions (kg)", fontsize=12)
17 plt.grid(axis='y', linestyle='--', alpha=0.6)
18
19 # Show the chart
20 plt.show()
21
```

(1) Spark Jobs

E-Commerce Carbon Footprint

community.cloud.databricks.com/editor/notebooks/1066710222480676?o=4051678597146965

databricks

New

Workspace

Recents

Search

Catalog

Workflows

Compute

Machine Learning

Experiments

E-Commerce Carbon Footprint Analytics. Python

File Edit View Run Help Last edit was 6 hours ago

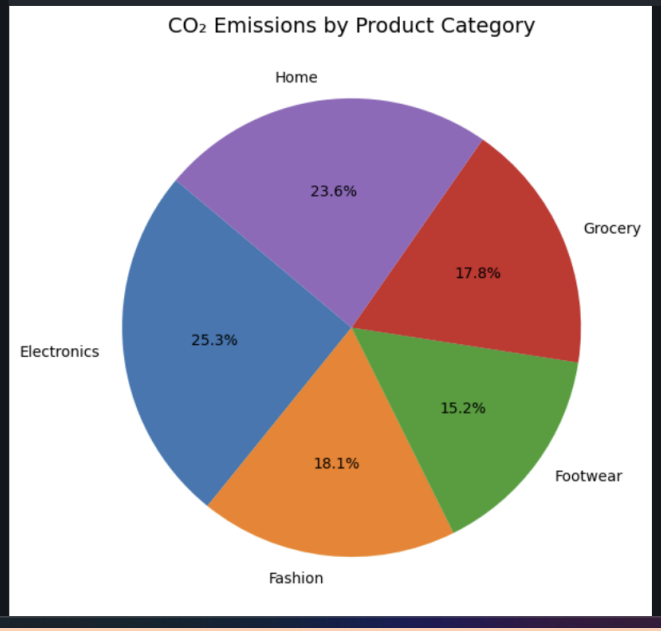
Run all carbon-footprint-cluster Share Publish

Pie chart –CO_2 emissions by product category

12:00 PM (<1s) 25

```
# Convert Carbon_Emission_kg to numeric
pdf["Carbon_Emission_kg"] = pd.to_numeric(pdf["Carbon_Emission_kg"], errors='coerce')
pdf = pdf.dropna(subset=["Carbon_Emission_kg"])
category_summary = pdf.groupby("Category")["Carbon_Emission_kg"].sum().reset_index()

plt.figure(figsize=(7, 7))
plt.pie(category_summary["Carbon_Emission_kg"], labels=category_summary["Category"], autopct='%1.1f%%', startangle=140)
plt.title("CO₂ Emissions by Product Category", fontsize=14)
plt.show()
```



E-Commerce Carbon Footprint

community.cloud.databricks.com/editor/notebooks/1066710222480676?o=4051678597146965

databricks

New

Workspace

Recents

Search

Catalog

Workflows

Compute

Machine Learning

Experiments

E-Commerce Carbon Footprint Analytics. Python

File Edit View Run Help Last edit was 6 hours ago

Run all carbon-footprint-cluster Share Publish

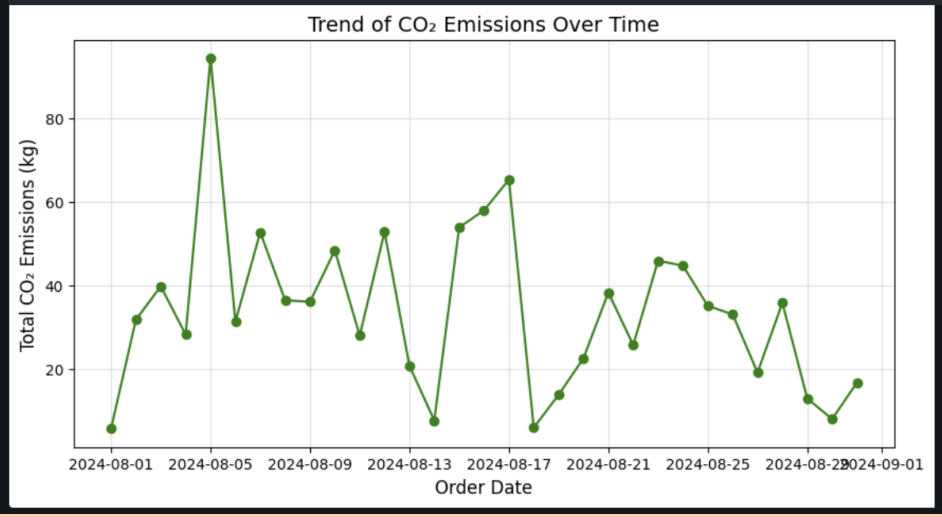
Line chart – Emission trends over time

12:01 PM (<1s) 27 Python

```
# Convert OrderDate to datetime
pdf["OrderDate"] = pd.to_datetime(pdf["OrderDate"])

# Group by OrderDate
date_summary = pdf.groupby("OrderDate")["Carbon_Emission_kg"].sum().reset_index()

plt.figure(figsize=(10, 5))
plt.plot(date_summary["OrderDate"], date_summary["Carbon_Emission_kg"], marker='o', color='green')
plt.title("Trend of CO₂ Emissions Over Time", fontsize=14)
plt.xlabel("Order Date", fontsize=12)
plt.ylabel("Total CO₂ Emissions (kg)", fontsize=12)
plt.grid(alpha=0.4)
plt.show()
```



E-Commerce Carbon Footprint

community.cloud.databricks.com/editor/notebooks/1066710222480676?o=4051678597146965

databricks

New

Workspace

Recents

Search

Catalog

Workflows

Compute

Machine Learning

Experiments

E-Commerce Carbon Footprint Analytics. Python

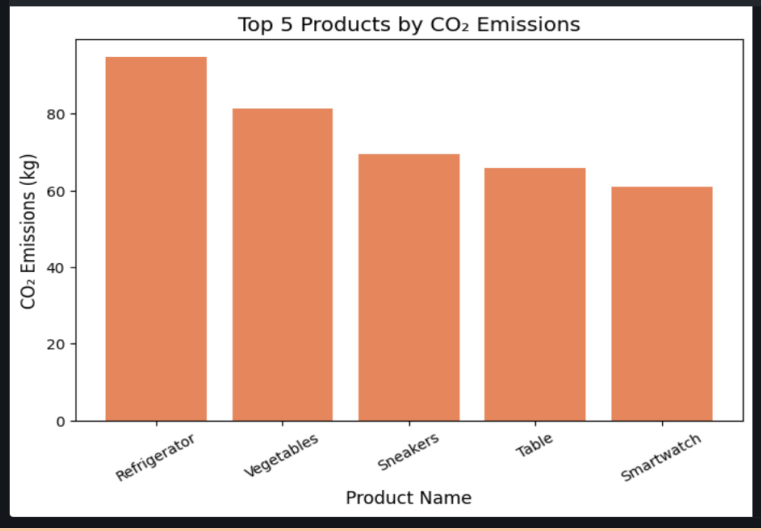
File Edit View Run Help Last edit was 6 hours ago

Run all carbon-footprint-cluster Share Publish

Top 5 products – Ranked by emissions

12:03 PM (<1s) 29

```
1  top_products = pdf.groupby("ProductName")["Carbon_Emission_kg"].sum().reset_index().sort_values(by="Carbon_Emission_kg", ascending=False).head(5)
2
3  plt.figure(figsize=(8, 5))
4  plt.bar(top_products["ProductName"], top_products["Carbon_Emission_kg"], color='coral')
5  plt.title("Top 5 Products by CO₂ Emissions", fontsize=14)
6  plt.xlabel("Product Name", fontsize=12)
7  plt.ylabel("CO₂ Emissions (kg)", fontsize=12)
8  plt.xticks(rotation=30)
9  plt.show()
```



E-Commerce Carbon Footprint

community.cloud.databricks.com/editor/notebooks/1066710222480676?o=4051678597146965#command/1066710222480716

databricks

E-Commerce Carbon Footprint Analytics. Python

File Edit View Run Help Last edit was 6 hours ago

Run all | carbon-footprint-cluster | Share | Publish

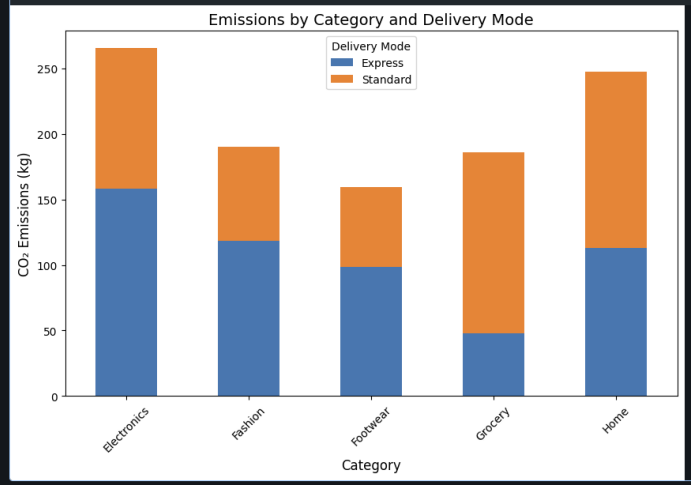
New
Workspace
Recents
Search
Catalog
Workflows
Compute
Machine Learning
Experiments

Stacked Bar: Category vs Delivery Mode

12:04 PM (<1s) 31 Python

```
pivot_data = pdf.pivot_table(values="Carbon_Emission_kg", index="Category", columns="DeliveryMode", aggfunc="sum", fill_value=0)

pivot_data.plot(kind="bar", stacked=True, figsize=(10, 6))
plt.title("Emissions by Category and Delivery Mode", fontsize=14)
plt.xlabel("Category", fontsize=12)
plt.ylabel("CO₂ Emissions (kg)", fontsize=12)
plt.xticks(rotation=45)
plt.legend(title="Delivery Mode")
plt.show()
```



E-Commerce Carbon Footprint

community.cloud.databricks.com/editor/notebooks/1066710222480676?o=4051678597146965#command/1066710222480716

databricks

New

Workspace

Recents

Search

Catalog

Workflows

Compute

Machine Learning

Experiments

E-Commerce Carbon Footprint Analytics. Python

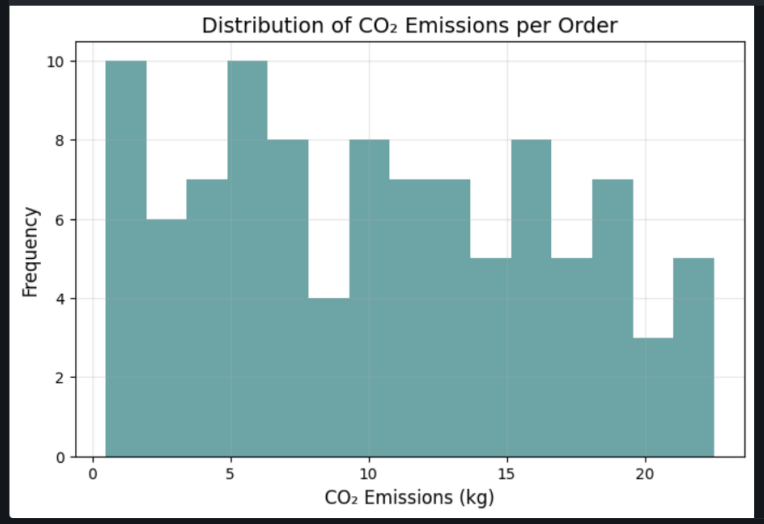
File Edit View Run Help Last edit was 6 hours ago

Run all carbon-footprint-cluster Share Publish

Histogram: Distribution of Emissions

12:05 PM (<1s) 33 Python

```
1  plt.figure(figsize=(8, 5))
2  plt.hist(pdf["Carbon_Emission_kg"], bins=15, color='teal', alpha=0.7)
3  plt.title("Distribution of CO₂ Emissions per Order", fontsize=14)
4  plt.xlabel("CO₂ Emissions (kg)", fontsize=12)
5  plt.ylabel("Frequency", fontsize=12)
6  plt.grid(alpha=0.3)
7  plt.show()
8
```



E-Commerce Carbon Footprint

community.cloud.databricks.com/editor/notebooks/1066710222480676?o=4051678597146965#command/1066710222480716

databricks

New

Workspace

Recents

Search

Catalog

Workflows

Compute

Machine Learning

Experiments

E-Commerce Carbon Footprint Analytics. Python

File Edit View Run Help Last edit was 6 hours ago

Run all carbon-footprint-cluster Share Publish

Scatter Plot: Distance vs Emissions

12:05 PM (1s) 35 Python

```
plt.figure(figsize=(8, 5))
plt.scatter(pdf["Distance_km"], pdf["Carbon_Emission_kg"], color='orange', alpha=0.6)
plt.title("Distance vs CO₂ Emissions", fontsize=14)
plt.xlabel("Distance (km)", fontsize=12)
plt.ylabel("CO₂ Emissions (kg)", fontsize=12)
plt.grid(alpha=0.3)
plt.show()
```