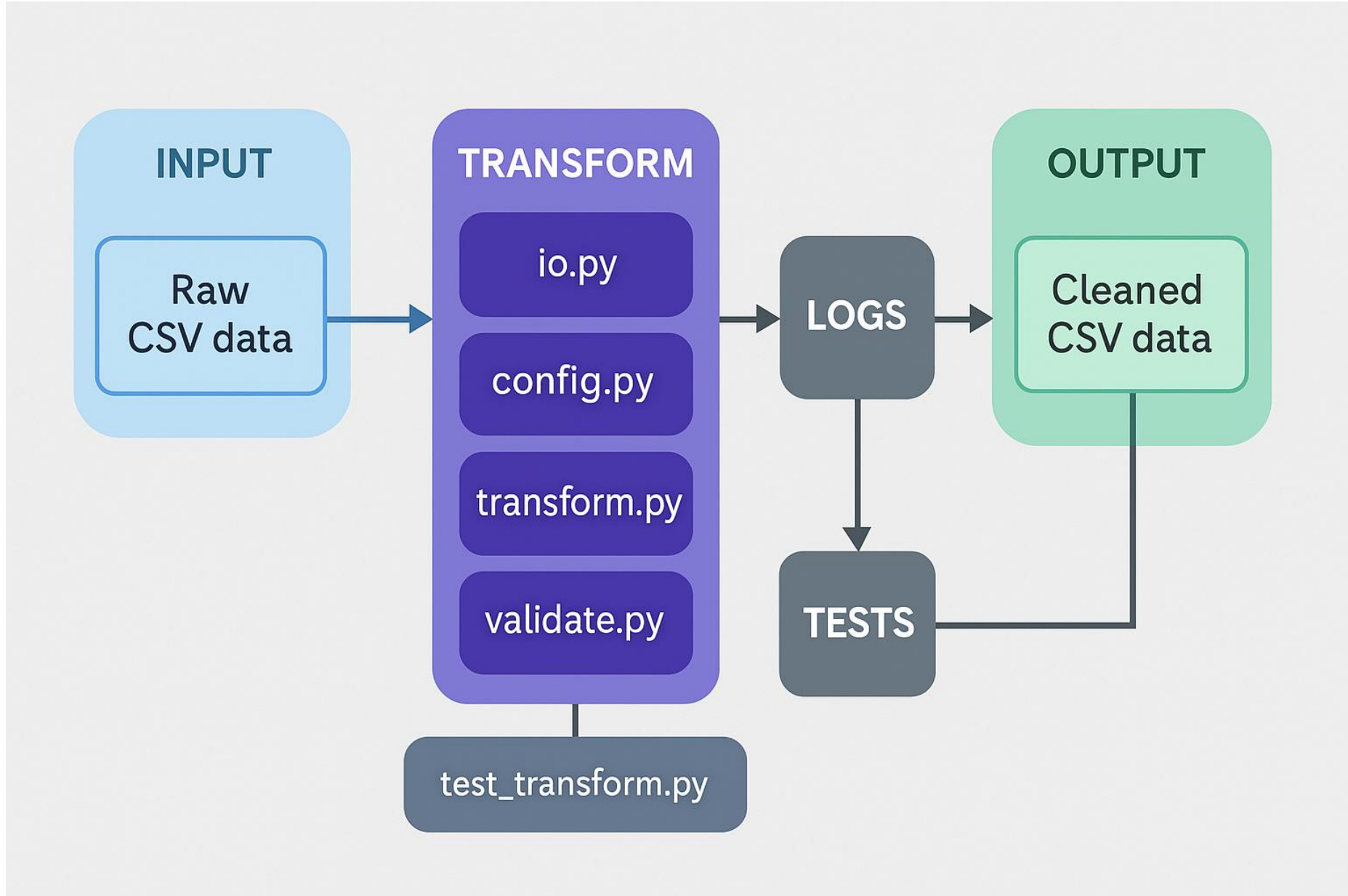
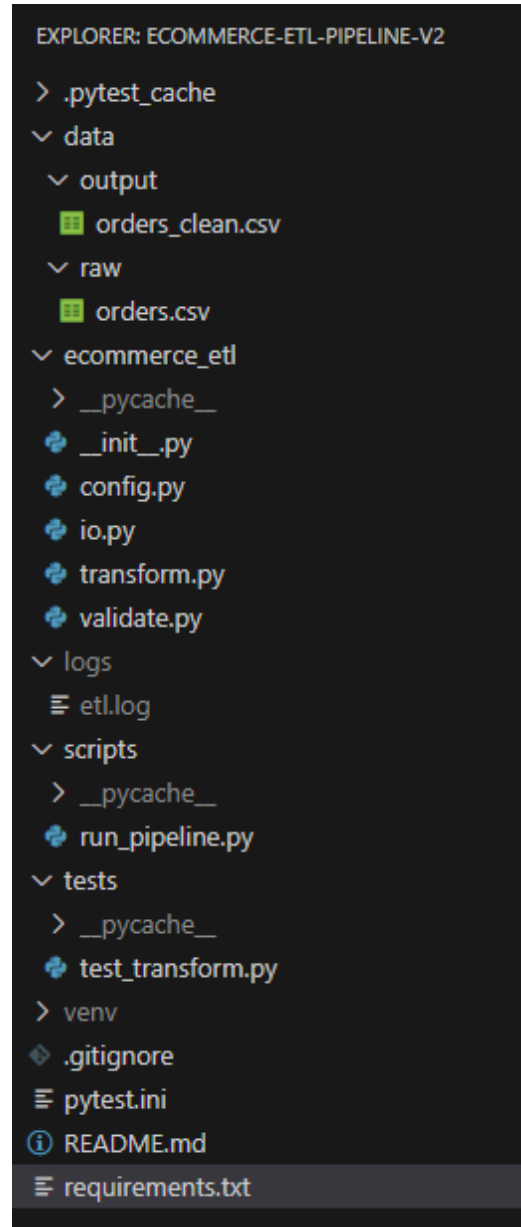


Data Flow Structure



Folders Structure



Data Raw

```
data > raw > orders.csv > data
```



```
1  order_id,order_date,customer_id,product_id,product_name,quantity,unit_price,currency,status
2  1001,1/15/2025,501,78,Wireless Mouse,2,19.99,USD,shipped
3  1002,1/16/2025,502,79,Gaming Keyboard,1,49.99,USD,pending
4  1003,1/17/2025,503,80,USB-C Hub,3,29.99,USD,returned
5  1004,1/17/2025,504,81,Laptop Stand,1,39.99,USD,shipped
6  1005,1/18/2025,505,82,Noise Cancelling Headphones,1,89.99,USD,cancelled
7  1006,1/18/2025,506,83,Webcam HD,2,45.99,USD,shipped
8  1007,1/19/2025,507,84,Portable SSD 1TB,1,129.99,USD,pending
9  1008,1/19/2025,508,85,Ergonomic Chair,1,199.99,USD,shipped
10 1009,1/20/2025,509,86,Monitor 24",2,139.99,USD,shipped
11 1010,1/20/2025,510,87,Wireless Charger,4,24.99,USD,returned
12 1011,1/21/2025,511,88,Mechanical Keyboard,1,79.99,USD,shipped
13 1012,1/21/2025,512,89,Stylus Pen,2,15.99,USD,pending
14 1013,1/22/2025,513,90,Laptop Backpack,1,59.99,USD,shipped
15 1014,1/22/2025,514,91,HDMI Cable,3,9.99,USD,cancelled
16 1015,1/23/2025,515,92,Graphic Tablet,1,149.99,USD,shipped
17 1016,1/23/2025,516,93,Smartwatch,2,199.99,USD,pending
18 1017,1/24/2025,517,94,VR Headset,1,299.99,USD,shipped
19 1018,1/24/2025,518,95,Desk Lamp LED,2,34.99,USD,shipped
20 1019,1/25/2025,519,96,Smartphone Stand,3,11.99,USD,returned
21 1020,1/25/2025,520,97,USB Drive 64GB,5,12.99,USD,shipped
22 1021,1/26/2025,521,98,Bluetooth Speaker,1,59.99,USD,cancelled
23 1022,1/26/2025,522,99,Wireless Earbuds,2,89.99,USD,shipped
24 1023,1/27/2025,523,100,Router Wi-Fi 6,1,129.99,USD,shipped
25 1024,1/27/2025,524,101,Laptop Cooling Pad,1,24.99,USD,pending
26 1025,1/28/2025,525,102,External Hard Drive 2TB,1,89.99,USD,shipped
27 1026,1/28/2025,526,103,USB-C to HDMI Adapter,3,19.99,USD,returned
28 1027,1/29/2025,527,104,Compact Bluetooth Mouse,2,25.99,USD,shipped
29 1028,1/29/2025,528,105,Noise Isolating Earbuds,1,39.99,USD,pending
30 1029,1/30/2025,529,106,Power Bank 10000mAh,2,29.99,USD,shipped
31 1030,1/30/2025,530,107,Curved Monitor 27",1,239.99,USD,shipped
```

Requirements.txt


≡ requirements.txt

```
1  numpy==2.3.1
2  pandas==2.3.1
3  python-dateutil==2.9.0.post0
4  pytz==2025.2
5  six==1.17.0
6  tzdata==2025.2
7  pytest
8
```

io.py

```
ecommerce_etl >  io.py >  write_csv  
1  import pandas as pd  
2  
3  def read_csv(path):  
4      |    return pd.read_csv(path)  
5  
6  def write_csv(df, path):  
7      |    df.to_csv(path, index=False)
```

Config.py

```
ecommerce_etl >  config.py > ...
```

```
1  from pathlib import Path
```

```
2
```

```
3  BASE_DIR = Path(__file__).resolve().parent.parent
```

```
4  RAW_PATH = BASE_DIR / "data" / "raw" / "orders.csv"
```

```
5  OUTPUT_PATH = BASE_DIR / "data" / "output" / "orders_clean.csv"
```

```
6  LOG_PATH = BASE_DIR / "logs" / "etl.log"
```

Transform.py

ecommerce_etl > transform.py > ...

```
1  import pandas as pd
2
3  def standardize_column_names(df):
4      df.columns = (
5          df.columns
6          .str.strip()
7          .str.lower()
8          .str.replace(" ", "_")
9      )
10     return df
11
12     def convert_types(df):
13         df["order_id"] = df["order_id"].astype("int64")
14         df["customer_id"] = df["customer_id"].astype("int64")
15         df["product_id"] = df["product_id"].astype("int64")
16         df["quantity"] = df["quantity"].astype("int64")
17         df["unit_price"] = df["unit_price"].astype("float64")
18         df["order_date"] = pd.to_datetime(df["order_date"], errors="coerce")
19         return df
20
21     def add_total_amount(df):
22         df["total_amount"] = df["quantity"] * df["unit_price"]
23         return df
24
```

Validate.py

ecommerce_etl > validate.py > ...

```
1  def no_nulls(df, cols):
2      for col in cols:
3          if df[col].isna().any():
4              raise ValueError(f"{col} contiene NULLs")
5
6  def positive_values(df, cols):
7      for col in cols:
8          if (df[col] < 0).any():
9              raise ValueError(f"{col} contiene negativos")
10
11  def allowed_status(df):
12      allowed = {"pending", "shipped", "returned", "cancelled"}
13      bad = df.loc[~df["status"].isin(allowed), "status"].unique()
14      if len(bad):
15          raise ValueError(f"status inválido: {bad}")
16
```


Test_transform.py

tests > test_transform.py > ...

```
4 def test_total_amount():
5     df = pd.DataFrame({
6         "quantity": [2, 3],
7         "unit_price": [10.0, 5.0]
8     })
9
10    df = add_total_amount(df)
11    assert df["total_amount"].tolist() == [20.0, 15.0]
12
```

Etl.log

logs > ≡ etl.log

1	2025-07-09 17:41:39,042		INFO		Inicio del pipeline
2	2025-07-09 17:41:39,057		INFO		Fin del pipeline
3	2025-07-20 16:48:20,697		INFO		Inicio del pipeline
4	2025-07-20 16:48:20,841		INFO		Fin del pipeline
5	2025-07-20 16:55:16,028		INFO		Inicio del pipeline
6	2025-07-20 16:55:16,044		INFO		Fin del pipeline
7					

Data Cleaned Output

```
data > output > orders_clean.csv > data
```

```
1  order_id,order_date,customer_id,product_id,product_name,quantity,unit_price,currency,status,total_amount
2  1001,2025-01-15,501,78,Wireless Mouse,2,19.99,USD,shipped,39.98
3  1002,2025-01-16,502,79,Gaming Keyboard,1,49.99,USD,pending,49.99
4  1003,2025-01-17,503,80,USB-C Hub,3,29.99,USD,returned,89.97
5  1004,2025-01-17,504,81,Laptop Stand,1,39.99,USD,shipped,39.99
6  1005,2025-01-18,505,82,Noise Cancelling Headphones,1,89.99,USD,cancelled,89.99
7  1006,2025-01-18,506,83,Webcam HD,2,45.99,USD,shipped,91.98
8  1007,2025-01-19,507,84,Portable SSD 1TB,1,129.99,USD,pending,129.99
9  1008,2025-01-19,508,85,Ergonomic Chair,1,199.99,USD,shipped,199.99
10 1009,2025-01-20,509,86,"Monitor 24\"",2,139.99,USD,shipped,279.98
11 1010,2025-01-20,510,87,Wireless Charger,4,24.99,USD,returned,99.96
12 1011,2025-01-21,511,88,Mechanical Keyboard,1,79.99,USD,shipped,79.99
13 1012,2025-01-21,512,89,Stylus Pen,2,15.99,USD,pending,31.98
14 1013,2025-01-22,513,90,Laptop Backpack,1,59.99,USD,shipped,59.99
15 1014,2025-01-22,514,91,HDMI Cable,3,9.99,USD,cancelled,29.97
16 1015,2025-01-23,515,92,Graphic Tablet,1,149.99,USD,shipped,149.99
17 1016,2025-01-23,516,93,Smartwatch,2,199.99,USD,pending,399.98
18 1017,2025-01-24,517,94,VR Headset,1,299.99,USD,shipped,299.99
19 1018,2025-01-24,518,95,Desk Lamp LED,2,34.99,USD,shipped,69.98
20 1019,2025-01-25,519,96,Smartphone Stand,3,11.99,USD,returned,35.97
21 1020,2025-01-25,520,97,USB Drive 64GB,5,12.99,USD,shipped,64.95
22 1021,2025-01-26,521,98,Bluetooth Speaker,1,59.99,USD,cancelled,59.99
23 1022,2025-01-26,522,99,Wireless Earbuds,2,89.99,USD,shipped,179.98
24 1023,2025-01-27,523,100,Router Wi-Fi 6,1,129.99,USD,shipped,129.99
25 1024,2025-01-27,524,101,Laptop Cooling Pad,1,24.99,USD,pending,24.99
26 1025,2025-01-28,525,102,External Hard Drive 2TB,1,89.99,USD,shipped,89.99
27 1026,2025-01-28,526,103,USB-C to HDMI Adapter,3,19.99,USD,returned,59.97
28 1027,2025-01-29,527,104,Compact Bluetooth Mouse,2,25.99,USD,shipped,51.98
29 1028,2025-01-29,528,105,Noise Isolating Earbuds,1,39.99,USD,pending,39.99
30 1029,2025-01-30,529,106,Power Bank 10000mAh,2,29.99,USD,shipped,59.98
31 1030,2025-01-30,530,107,"Curved Monitor 27\"",1,239.99,USD,shipped,239.99
32
```