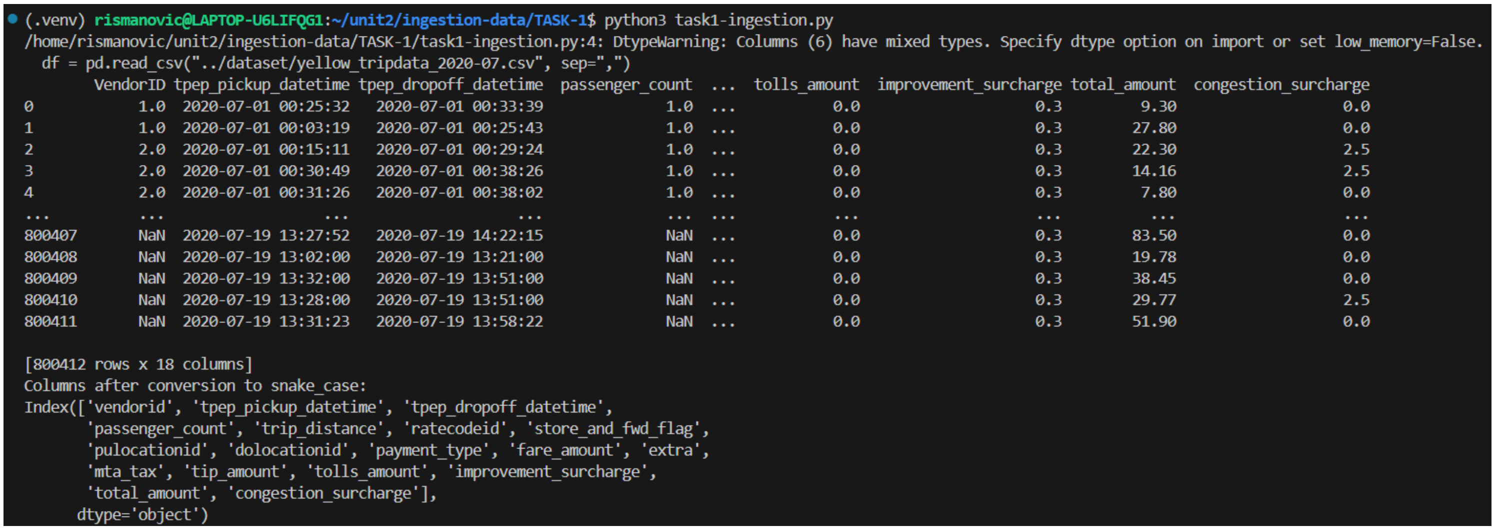
Nama : Kharisma Novi Chandramukti

Kelas/Batch : Data Engineering 4

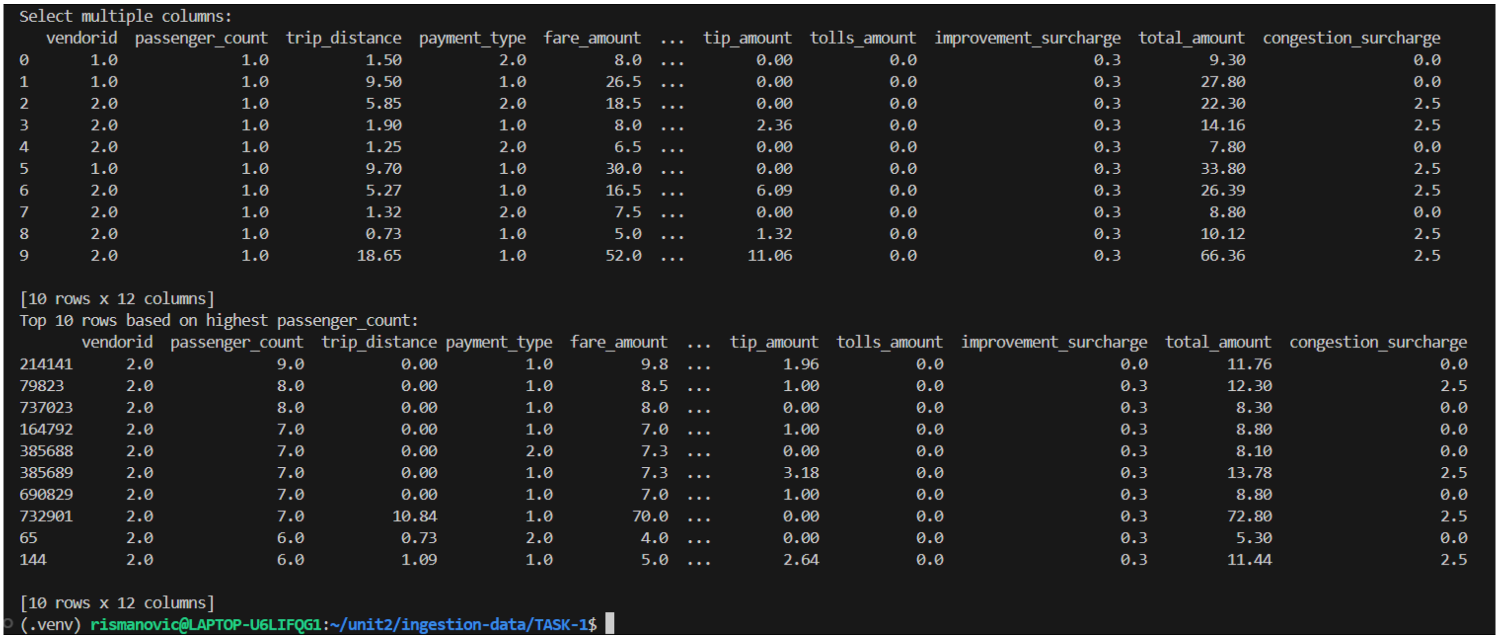
Task 1

Screenshot output pada Task 1



```
● (.venv) rismanovic@LAPTOP-U6LIFQG1:~/unit2/ingestion-data/TASK-1$ python3 task1-ingestion.py
  /home/rismanovic/unit2/ingestion-data/TASK-1/task1-ingestion.py:4: DtypeWarning: Columns (6) have mixed types. Specify dtype option on import or set low_memory=False.
    df = pd.read_csv("../dataset/yellow_tripdata_2020-07.csv", sep=",")
          VendorID tpep_pickup_datetime tpep_dropoff_datetime  passenger_count  ...  tolls_amount  improvement_surcharge total_amount  congestion_surcharge
  0            1.0  2020-07-01 00:25:32   2020-07-01 00:33:39              1.0  ...           0.0                    0.3         9.30                   0.0
  1            1.0  2020-07-01 00:03:19   2020-07-01 00:25:43              1.0  ...           0.0                    0.3        27.80                   0.0
  2            2.0  2020-07-01 00:15:11   2020-07-01 00:29:24              1.0  ...           0.0                    0.3        22.30                   2.5
  3            2.0  2020-07-01 00:30:49   2020-07-01 00:38:26              1.0  ...           0.0                    0.3        14.16                   2.5
  4            2.0  2020-07-01 00:31:26   2020-07-01 00:38:02              1.0  ...           0.0                    0.3         7.80                   0.0
  ...          ...                  ...                   ...              ...  ...           ...                    ...          ...                   ...
  800407       NaN  2020-07-19 13:27:52   2020-07-19 14:22:15              NaN  ...           0.0                    0.3        83.50                   0.0
  800408       NaN  2020-07-19 13:02:00   2020-07-19 13:21:00              NaN  ...           0.0                    0.3        19.78                   0.0
  800409       NaN  2020-07-19 13:32:00   2020-07-19 13:51:00              NaN  ...           0.0                    0.3        38.45                   0.0
  800410       NaN  2020-07-19 13:28:00   2020-07-19 13:51:00              NaN  ...           0.0                    0.3        29.77                   2.5
  800411       NaN  2020-07-19 13:31:23   2020-07-19 13:58:22              NaN  ...           0.0                    0.3        51.90                   0.0

  [800412 rows x 18 columns]
  Columns after conversion to snake_case:
  Index(['vendorid', 'tpep_pickup_datetime', 'tpep_dropoff_datetime',
         'passenger_count', 'trip_distance', 'ratecodeid', 'store_and_fwd_flag',
         'pulocationid', 'dolocationid', 'payment_type', 'fare_amount', 'extra',
         'mta_tax', 'tip_amount', 'tolls_amount', 'improvement_surcharge',
         'total_amount', 'congestion_surcharge'],
        dtype='object')
```



```
Select multiple columns:
   vendorid  passenger_count  trip_distance  payment_type  fare_amount  ...  tip_amount  tolls_amount  improvement_surcharge  total_amount  congestion_surcharge
0       1.0              1.0           1.50           2.0          8.0  ...        0.00           0.0                    0.3          9.30                   0.0
1       1.0              1.0           9.50           1.0         26.5  ...        0.00           0.0                    0.3         27.80                   0.0
2       2.0              1.0           5.85           2.0         18.5  ...        0.00           0.0                    0.3         22.30                   2.5
3       2.0              1.0           1.90           1.0          8.0  ...        2.36           0.0                    0.3         14.16                   2.5
4       2.0              1.0           1.25           2.0          6.5  ...        0.00           0.0                    0.3          7.80                   0.0
5       1.0              1.0           9.70           1.0         30.0  ...        0.00           0.0                    0.3         33.80                   2.5
6       2.0              1.0           5.27           1.0         16.5  ...        6.09           0.0                    0.3         26.39                   2.5
7       2.0              1.0           1.32           2.0          7.5  ...        0.00           0.0                    0.3          8.80                   0.0
8       2.0              1.0           0.73           1.0          5.0  ...        1.32           0.0                    0.3         10.12                   2.5
9       2.0              1.0          18.65           1.0         52.0  ...       11.06           0.0                    0.3         66.36                   2.5

[10 rows x 12 columns]
Top 10 rows based on highest passenger_count:
        vendorid  passenger_count  trip_distance payment_type  fare_amount  ...  tip_amount  tolls_amount  improvement_surcharge  total_amount  congestion_surcharge
214141       2.0              9.0           0.00          1.0          9.8  ...        1.96           0.0                    0.0         11.76                   0.0
79823        2.0              8.0           0.00          1.0          8.5  ...        1.00           0.0                    0.3         12.30                   2.5
737023       2.0              8.0           0.00          1.0          8.0  ...        0.00           0.0                    0.3          8.30                   0.0
164792       2.0              7.0           0.00          1.0          7.0  ...        1.00           0.0                    0.3          8.80                   0.0
385688       2.0              7.0           0.00          2.0          7.3  ...        0.00           0.0                    0.3          8.10                   0.0
385689       2.0              7.0           0.00          1.0          7.3  ...        3.18           0.0                    0.3         13.78                   2.5
690829       2.0              7.0           0.00          1.0          7.0  ...        1.00           0.0                    0.3          8.80                   0.0
732901       2.0              7.0          10.84          1.0         70.0  ...        0.00           0.0                    0.3         72.80                   2.5
65           2.0              6.0           0.73          2.0          4.0  ...        0.00           0.0                    0.3          5.30                   0.0
144          2.0              6.0           1.09          1.0          5.0  ...        2.64           0.0                    0.3         11.44                   2.5

[10 rows x 12 columns]
(.venv) rismanovic@LAPTOP-U6LIFQG1:~/unit2/ingestion-data/TASK-1$
```