```python
#!/bin/env python3

#
#
https://github.com/lloydalexporter/UChicago_MS_ADS_Supplem
ent
#

#!> Import libraries
import pandas as pd
import matplotlib.pyplot as plt
import numpy as np
from time import sleep
from pprint import pprint


# *** CONSTANTS
CSV_FILE_PATH = "weather.csv" # Path to the input CSV file
https://corgis-edu.github.io/corgis/csv/weather/
STATION_CODE = "ORD"          # Weather station code for
Chicago ;)



# !> Visualise data
def graph_precipitation(data):
  plt.figure(figsize=(10,5))
  plt.plot(data['Date.Full'], data['Data.Precipitation'],
marker='o', linestyle='-')
  plt.title(f'Precipitation in {STATION_CODE}')
```

```
  plt.xlabel('Date')
  plt.ylabel('Precipitation (Inches)')
  plt.grid(True)
  plt.show()

# !> Use your function for data analysis
def analyse_station_data(station_data):
  print(f"\nAnalysis for Station: {STATION_CODE}")

  num_days = len(station_data)
  total_precipitation =
station_data['Data.Precipitation'].sum()
  avg_precipitation =
station_data['Data.Precipitation'].mean()
  max_precipitation =
station_data['Data.Precipitation'].max()
  min_precipitation =
station_data['Data.Precipitation'].min()

  avg_temp = station_data['Data.Temperature.Avg
Temp'].mean()
  max_temp = station_data['Data.Temperature.Max
Temp'].max()
  min_temp = station_data['Data.Temperature.Min
Temp'].min()

  print(f"Total Precipitation: {total_precipitation:.2f}
inches over {num_days} days")
  print(f"Max ({max_precipitation:.2f}) | Avg
({avg_precipitation:.2f}) | Min ({min_precipitation:.2f})
Precipitation")
```

```
  print("Temperature Analysis (*F):")
  print(f"Max ({max_temp:.2f}) | Avg ({avg_temp:.2f}) |
Min ({min_temp:.2f})")


def preview_table(data, num_rows=5):
  print(f'\n' * 3)
  pprint(data.head(num_rows))
  print(data.dtypes)


# !> Manage different data types
def process_csv(raw_data):
  # Format the dates and separate into components
  raw_data["Date.Full"] =
pd.to_datetime(raw_data["Date.Full"])
  raw_data["Date.Month"] = raw_data["Date.Full"].dt.month
  raw_data["Date.Day"] = raw_data["Date.Full"].dt.day
  raw_data["Date.Year"] = raw_data["Date.Full"].dt.year

  return raw_data


def main():

  # !> Ingest data from CSV
  raw_data = pd.read_csv(CSV_FILE_PATH)
  headers = raw_data.columns.tolist()
  preview_table(raw_data)

  # !> Manage different data types
  processed_data = process_csv(raw_data)
```

```
  processed_data = process_csv(raw_data)
  preview_table(processed_data)

  # !> Wrangle data
  station_data =
processed_data[processed_data['Station.Code'] ==
STATION_CODE].copy()
  group_columns =
['Date.Full','Station.Code','Station.Location']
  aggregate = {
    'Data.Precipitation': 'sum',
    'Data.Temperature.Avg Temp': 'mean',
    'Data.Temperature.Max Temp': 'max',
    'Data.Temperature.Min Temp': 'min'
  }
  station_data =
station_data.groupby(group_columns).agg(aggregate).reset_i
ndex()

  preview_table(station_data)
  analyse_station_data(station_data)
  graph_precipitation(station_data)


if __name__ == "__main__":
  main()
```