#data wrangling with pandas

import pandas as pd

# Loading CSV files into DataFrames

df1 = pd.read\_csv("data/mexico-real-estate-1.csv")

df2 = pd.read\_csv("data/mexico-real-estate-2.csv")

df3 = pd.read\_csv("data/mexico-real-estate-3.csv")

# Drop null values from df1

df1.dropna(inplace=True)

df1.info()

# Clean "price\_usd" column in df1

df1.dropna( inplace=True)

df1 = df1.dropna(subset=["price\_usd"]) # Specify column to avoid dropping entire rows

# Remove characters and convert to float

df1["price\_usd"] = df1["price\_usd"].str.replace("$", "", regex=False).str.replace(",", "").astype(float)

<class 'pandas.core.frame.DataFrame'>

Int64Index: 583 entries, 0 to 699

Data columns (total 6 columns):

# Column Non-Null Count Dtype

--- ------ -------------- -----

0 property\_type 583 non-null object

1 state 583 non-null object

2 lat 583 non-null float64

3 lon 583 non-null float64

4 area\_m2 583 non-null float64

5 price\_usd 583 non-null object

dtypes: float64(3), object(3)

memory usage: 31.9+ KB

df1.head()

# Drop null values from df2

df2.dropna(inplace=True)

# Create "price\_usd" column for df2 (19 pesos to the dollar in 2014)

df2["price\_usd"] = ((df2["price\_mxn"]/19).round(2))

# Drop "price\_mxn" column from df2

df2.drop(columns=["price\_mxn"],inplace=True)

df2.head()

# Drop null values from df3

df3.dropna(inplace=True)

# Create "lat" and "lon" columns for df3

df3[["lat", "lon"]] = df3["lat-lon"].str.split(",", expand=True)

# Create "state" column for df3

df3["state"] = df3["place\_with\_parent\_names"].str.split("|",expand=True)[2]

# Drop "place\_with\_parent\_names" and "lat-lon" from df3

df3.drop(columns=["place\_with\_parent\_names","lat-lon"],inplace=True)

df3.head()

# Concatenate df1, df2, and df3

df = pd.concat([df1,df2,df3])

# Save df as a CSV

df.to\_csv("data/mexico-real-estate-clean.csv",index=False)