Homebuyers Guide to the Texas Market!!! Where should you end up?

Analysis Summary:

As part of this activity, "hotspots" in Texas is determined based on housing prices, COVID surveillance and public safety.

-- Main Questions: --

- Housing Price: Where are housing prices rising the most in the past 6 years? Where have prices fallen? AKA where do you want to stay away from in your home investment?
- COVID surveillance: Does the place have higher COVID transmission rate?
- Public safety: Is the place safe to live in? What is the crime ranking?

-- Data Analysis -- Data set for each question was identified, cleaned and a detailed summary analysis was performed.

-- Livability Score -- Based on the Data Analysis, from 1 to 5 with 1 being lowest and 5 being highest(best) is determined for each of these categories and overall livability grade based on the mean value of these scores is determined for each county

-- API Call -- Make an API call to get population count for all counties.

-- A visualization map -- Display Texas map by counties with a hover that display overall Livability Grade, County Name, Population, Current Housing Value, Average COVID Count, Average Offense County

and display in

Written Report:

-- Housing Analysis --housing_prices_bycounty.ipynb under Analysis folder

- Analysis shows that housing values in Texas have experienced a significant increase from 2017 to 2023, with no counties showing negative growth. During this period, the increase in housing values per county ranged from 10% to 117%. As of 2023, the average home value in Texas counties ranges from approximately $50k to $655k.
- To visualize the year-to-year changes in housing prices by county, created a boxplot of the average pricing changes. The plot shows a sharp rise in prices from 2020 to 2021, followed by a softening of the market since then.

-- COVID_Analysis_Summary -- COVID_Analysis.ipynb under Analysis folder

- Top Preferred 5 best counties based on COVID Analysis are "King, Shackelford, Foard, Stonewall and Runnels" whereas 5 least preferred counties are "Grimes, McLennan, Williamson, Cameron and Maverick."
- Also based on the COVID Spread Rate and COVID Avg Percent analysis, it is evident that "King County" has lowest covid count even though "Swisher County" has lowest COVID spread rate and it is evident that "Harris County" has highest covid count even though "Yoakum County" has highest COVID spread rate

-- SAFETY_Analysis_Summary -- SAFETY_Analysis_final.ipynb under Analysis folder

1. Based on analysis it is found Harris County has the total record of 8146 Offenses followed by Tarrant (5917 -Offense) and Bexar Counties (4317).
2. Year 2018 has the highest offense records. Most of them were intoxicated driving, sex offenders' duty to register and they are classified as other category.
3. Year 2020 has the lowest offense records. This can be primarily due to lack of information for that year.
4. Even though the county with the highest Offense rate is Harris County with a whopping 12.86 offense Percentage, Anderson County is least preferred based on livability index.
5. Even though the county with the smallest offense percent is Armstrong County with 0.01 percentage, Irion County is the best and top preferred county based on livability index.
6. Counties are given Safety Index based on their Offense Percent with 5 being the best and 1 being the worst.
7. Total of 42 counties gets the best rating (5) and 32 counties get the worst rating (1).

-- Based on the Livability Grade -- HomeBuyersGuide.ipynb It is evident from the livability grade that

- "Mason County" is the best overall county to live based on all factors (Housing, COVID and Safety Grade)

- "Midland County, San Patricio County and Nueces County" are least preferred county.
- Also, best preferred counties have way lesser population when compared to least preferred counties

---

Limitations

-- Test data Limitation --

- Housing Dataset: Based on housing dataset, data for only 211 counties out of 254 have complete all years (from 2017 to 2022) data available. Missing counties might result in biased grading.
- COVID Dataset: Considering COVID Start year is from 2020 march, complete full year dataset that can used for analysis and evaluation will be 2021 and 2022 years only.
- Safety Issues Dataset: Public safety dataset that was made available in Texas open data portal was only for 2019 and 2020 year only and based on further analysis within the available data, 2020 year had only 72 records. If the data set, made available public is incomplete it can impact this analysis.

-- Linear Regression and Correlation Analysis Limitation-- Based on the test data limitation mentioned above, a linear regression and correlation analysis will result in biased result and hence it is not performed between Housing, COVID and Safety datasets.

Dependencies and Setup

```python
In [1]:  # import required libiraries/packages
         import pandas as pd
         import matplotlib.pyplot as plt
         import hvplot.pandas
         import geopandas as gpd
         import requests
         from pprint import pprint
```

Loading Housing, COVID and Safety Analysis Dataset from output folder and calculate Livability Grade

```python
In [2]:  # Import datasets csv files -  Liviablity Indexed Datasets from each of the Analysis
         COVID_DATAF = pd.read_csv("output/LIndexForCOVID.csv")
         SAFTEY_DATADF = pd.read_csv("output/safety_analysis.csv")

         HOUSING_DATADF = pd.read_csv("output/housing_price_score.csv")
         # Based on Analysis column header and one CountyName column needs update. Formatting Datasheet matching to othe
         HOUSING_DATADF.rename({'CountyName':'COUNTY','Housing Value % Change': 'HOUSING_PCT_CHNG'}, axis = "columns", i
         HOUSING_DATADF.loc[HOUSING_DATADF[HOUSING_DATADF['COUNTY']=='De Witt'].index.values,['COUNTY']] = "DeWitt"

         # Merge all Dataframes
         HomeBuyerGuideDF = pd.merge((pd.merge(COVID_DATAF,HOUSING_DATADF,how='outer', on="COUNTY")),SAFTEY_DATADF,how='
         HomeBuyerGuideDF = HomeBuyerGuideDF.fillna(0)
         HomeBuyerGuideDF
         # Calculate Overall livability Grade
         HomeBuyerGuideDF["LIVABILITY_GRADE"] = HomeBuyerGuideDF[['Housing Score','COVID_LIV_SCORE','SAFETY_LINDEX']].me
```

Using API Call get the count of population for each county and merge with HomeBuyerGuideDF

```
In [3]:   from config import cdc_apptoken

          # Set the API base URL to retrive population data by county for state of Texas
          url = 'https://data.cdc.gov/resource/3nnm-4jni.json?state=Texas'

          # Set the query params - county and County population count is required
          params = {
              '$select': 'county,county_population',
              '$limit': 254
          }

          # Define an empty list to add each record retrived from CDC
          County_Pop_Data = []

          # Make a GET request to the API endpoint with the specified parameters
          CountyPopulationDataResponse = requests.get(url,params=params)

          # Create counters
          record_count = 0

          # Check if the request was successful
          if CountyPopulationDataResponse.status_code == 200:
              # Print the confirmed COVID-19 cases by county and date
              for eachCounty in CountyPopulationDataResponse.json():
                  county = eachCounty['county']
                  County_Population = eachCounty['county_population']

                  # Append the each county and population information into County_Pop_Data list
                  County_Pop_Data.append({"COUNTY": county,
                                          "Population": County_Population
                                         })
                  record_count = record_count + 1
                  # pprint(eachCounty)
          else:
              print(f"Error: {CountyPopulationDataResponse.status_code}")

          # Create dataframe from the output records retrived from CDC
          CountyPopulationDF = pd.DataFrame(County_Pop_Data)

          # Format the county column and merge with Housing dataframe
          CountyPopulationDF['COUNTY'] = CountyPopulationDF['COUNTY'].str.replace(" County","")
          HomeBuyerGuideDF = pd.merge(CountyPopulationDF,HomeBuyerGuideDF,how='outer', on="COUNTY")
          HomeBuyerGuideDF = HomeBuyerGuideDF.fillna(0)

          # Display Dataframe
          HomeBuyerGuideDF
```

Out[3]:

	COUNTY	Population	AVG_COVID_COUNT	AVG_COVID_PERCENT	COVID_SPREAD_PCT	COVID_AVG_PCT_RANK	COVID_SPREAD_R
0	Anderson	57735	202.916667	25.722118	-47.622028	0.696850	0.02
1	Andrews	18705	74.500000	9.443767	93.114754	0.460630	0.98
2	Angelina	86715	267.000000	33.845448	-37.854251	0.744094	0.09
3	Aransas	23510	147.750000	18.729082	21.763602	0.606299	0.78
4	Archer	8553	70.083333	8.883902	48.449040	0.444882	0.89
...	...	...	...	...	...	...	
249	Wood	45539	167.041667	21.174532	-9.729473	0.629921	0.39
250	Yoakum	8713	18.916667	2.397914	448.571429	0.212598	1.00
251	Young	18010	99.208333	12.575845	-25.950292	0.531496	0.18
252	Zapata	14179	56.208333	7.125080	2.248876	0.393701	0.56
253	Zavala	11840	107.583333	13.637476	4.272152	0.559055	0.61

254 rows × 23 columns

Top 5 Best Counties to live in based on Livability Grade

In [4]: `display(pd.DataFrame(HomeBuyerGuideDF.sort_values(by=['LIVABILITY_GRADE'],ascending=False)[['COUNTY','Populatio`

	COUNTY	Population	LIVABILITY_GRADE	COVID_LIV_SCORE	Housing Score	SAFETY_LINDEX
0	Mason	4274	4.50	4.5	5.0	4
1	Stephens	9366	4.17	4.5	5.0	3
2	Delta	5331	4.17	3.5	5.0	4
3	Presidio	6704	4.17	2.5	5.0	5
4	San Augustine	8237	4.00	4.0	4.0	4

Top 5 Least preferred county based on Livability Grade

In [5]: `display(pd.DataFrame(HomeBuyerGuideDF.sort_values(by=['LIVABILITY_GRADE'],ascending=True)[['COUNTY','Population`

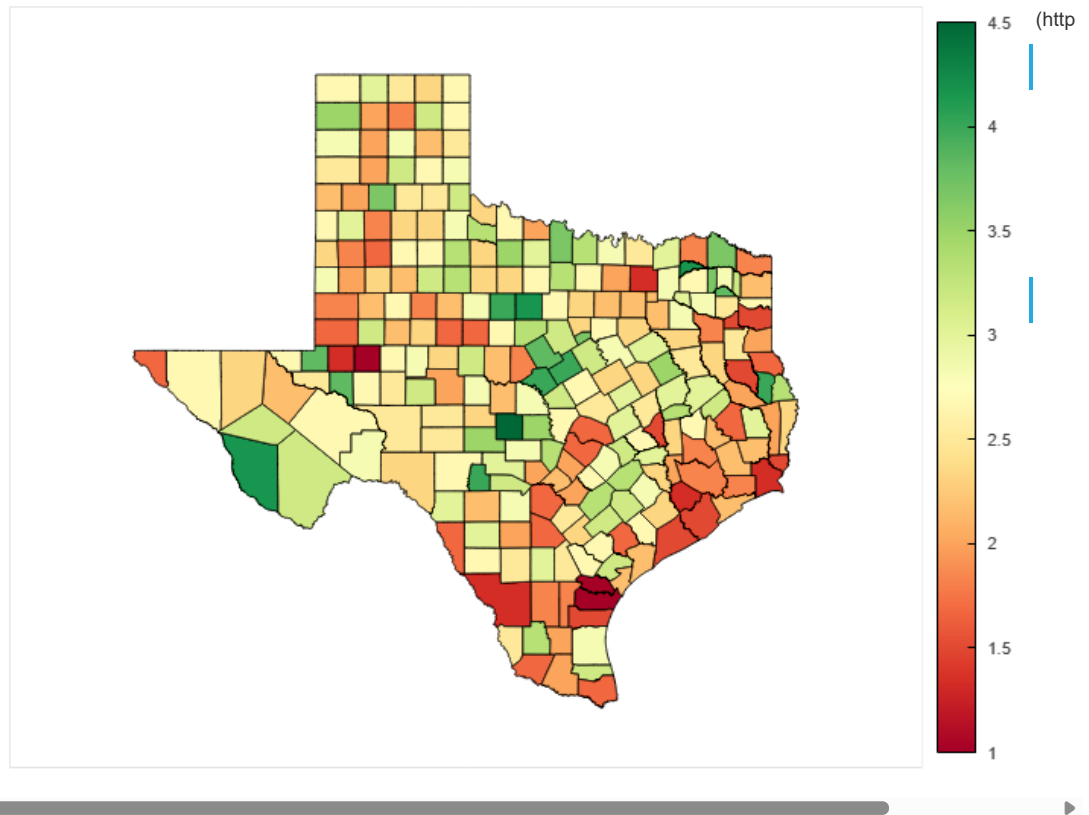|	COUNTY	Population	LIVABILITY_GRADE	COVID_LIV_SCORE	Housing Score	SAFETY_LINDEX
0	Midland	176832	1.00	1.0	1.0	1
1	San Patricio	66730	1.00	1.0	1.0	1
2	Nueces	362294	1.00	1.0	1.0	1
3	Webb	276652	1.33	1.0	2.0	1
4	Fort Bend	811688	1.33	1.0	2.0	1

Texas map visualization

In [6]:

```python
# Display Texas map by counties with a hover that display overall Livability Grade, County Name, Population, Cu

# Read the Texas counties list, merge with HomeBuyersGuideDF and create data frame for Texas Map
SHAPEFILE = 'Resources/tufts-txcounties10-shapefile/GISPORTAL_GISOWNER01_TXCOUNTIES10.shp'
Texas_Lat_Lng = gpd.read_file(SHAPEFILE)[['NAME10', 'INTPTLAT10','INTPTLON10','geometry']]
Texas_Lat_Lng.columns = ['COUNTY','LAT','LNG','geometry']
HomeBuyersGuideDF_Map = pd.merge(Texas_Lat_Lng, HomeBuyerGuideDF, on='COUNTY')

# Display Texas map with all required hover information
HomeBuyersGuideDF_Map.hvplot(c='LIVABILITY_GRADE',
                             frame_width=600, frame_height=500,
                             geo=True, color= 'LIVABILITY_GRADE', cmap='RdYlGn',
                             hover_cols=['COUNTY','Population','LIVABILITY_GRADE', 'Current Housing Value','AVG
```

Out[6]:

In [ ]: