Toronto Dwellings Analysis

In this assignment, you will perform fundamental analysis for the Toronto dwellings market to allow potential real estate investors to choose rental investment properties.

In [1]:
```
# Imports
import os
import pandas as pd
import matplotlib.pyplot as plt
import hvplot.pandas
import plotly.express as px
from pathlib import Path
from dotenv import load_dotenv
import matplotlib
import numpy as np

%matplotlib inline
```

In [2]:
```
# Read the Mapbox API key
load_dotenv()
map_box_api = os.getenv("MAP_BOX_API_KEY")

# Set the Mapbox API
px.set_mapbox_access_token(map_box_api)
```

Load Data

In [3]:
```
# Read the census data into a Pandas DataFrame
file_path = Path("Data/toronto_neighbourhoods_census_data.csv")
to_data = pd.read_csv(file_path, index_col="year")
to_data.head()
```

Out[3]:

year	neighbourhood	single_detached_house	apartment_five_storeys_plus	movable_dwelling	semi_d
2001	Agincourt North	3715	1480	0	
2001	Agincourt South-Malvern West	3250	1835	0	
2001	Alderwood	3175	315	0	
2001	Annex	1060	6090	5	
2001	Banbury-Don Mills	3615	4465	0	

Dwelling Types Per Year

In this section, you will calculate the number of dwelling types per year. Visualize the results using bar charts and the Pandas plot function.

Hint: Use the Pandas `groupby` function.

Optional challenge: Plot each bar chart in a different color.

In [4]:
```
# Calculate the sum number of dwelling types units per year (hint: use groupby)
dwelling_types = to_data.groupby('year').sum()
dwelling_types.head()
```

Out[4]:

year	single_detached_house	apartment_five_storeys_plus	movable_dwelling	semi_detached_house	r
2001	300930	355015	75	90995	
2006	266860	379400	165	69430	
2011	274940	429220	100	72480	
2016	269680	493270	95	71200	

In [5]:
```
# Drop unnecessary columns
dwelling_types = dwelling_types.drop(['average_house_value', 'shelter_costs_owned',
```

In [6]:
```
# Inspect DataFrame
dwelling_types.head()
```

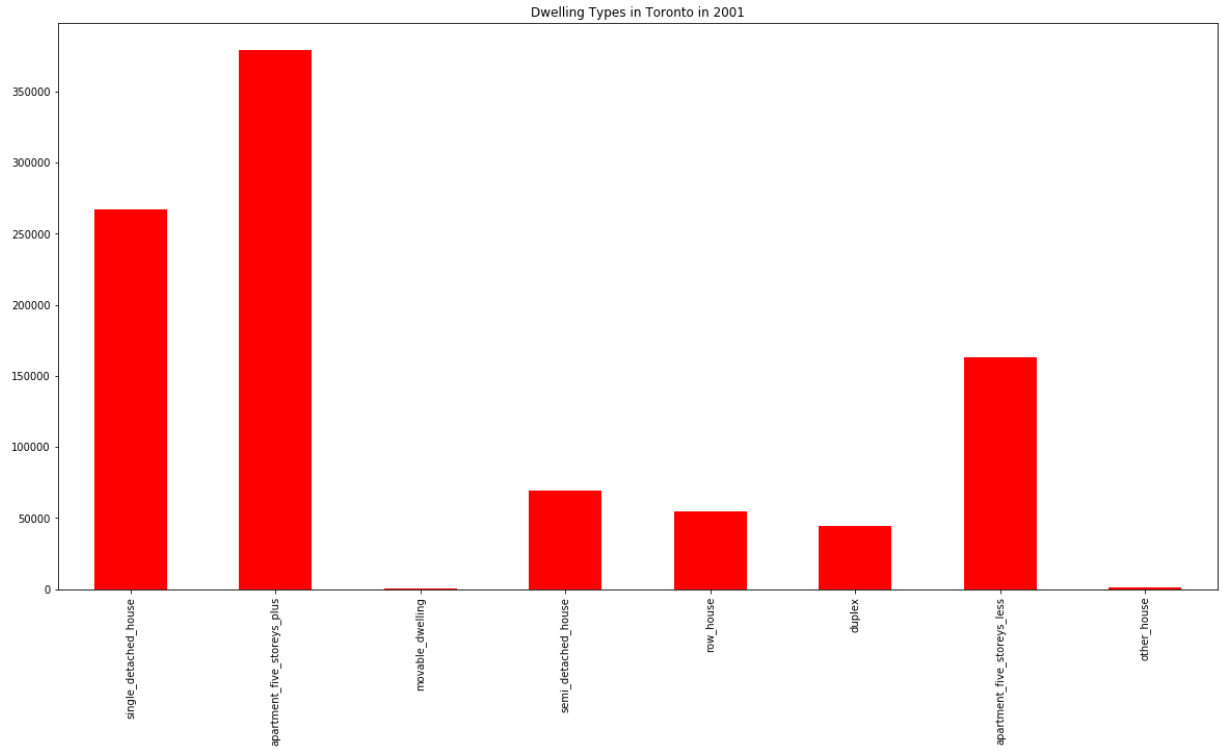
Out[6]:

year	single_detached_house	apartment_five_storeys_plus	movable_dwelling	semi_detached_house	r
2001	300930	355015	75	90995	
2006	266860	379400	165	69430	
2011	274940	429220	100	72480	
2016	269680	493270	95	71200	

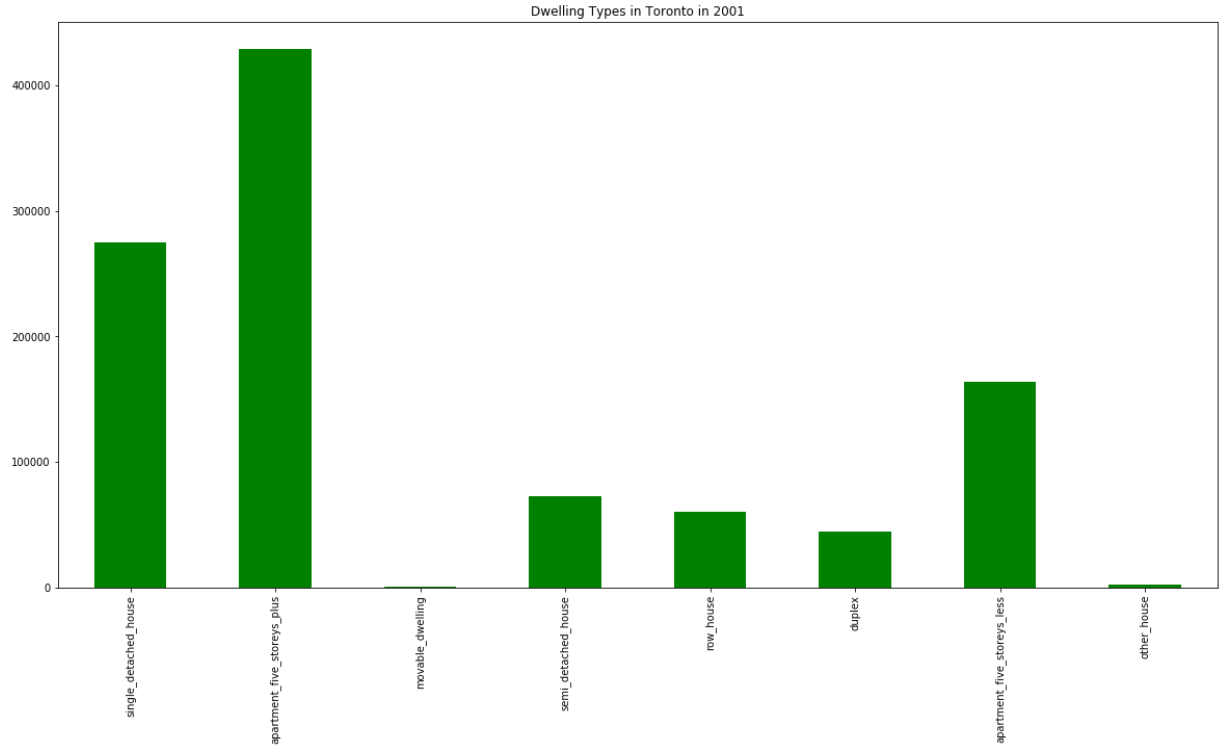
In [7]:
```
# Save the dataframe as a csv file
dwelling_types.to_csv =("dwelling_types_by_year.csv")
```

In [8]:
```
# Helper create_bar_chart function
def create_bar_chart(year, data, xlabel, ylabel, color):
    fig = plt.figure(figsize=(20,10))
    data.plot.bar(title = f'Dwelling Types in Toronto in {year}',
                  color = color)
    plt.show()
```

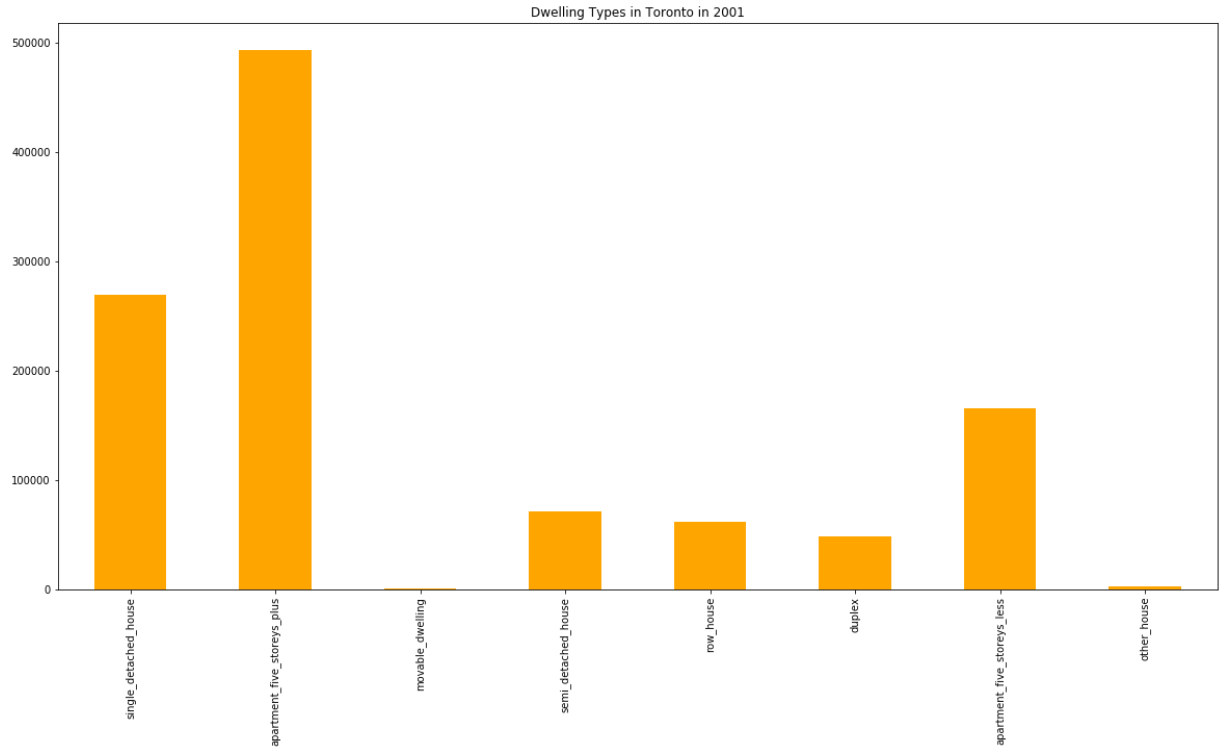
In [9]:
```
# Dwelling types - 2006 (Red)
create_bar_chart('2001', dwelling_types.iloc[1],'','', 'red')
```



In [10]:
```
# Dwelling types - 2011 (Red)
create_bar_chart('2001', dwelling_types.iloc[2],'','', 'green')
```

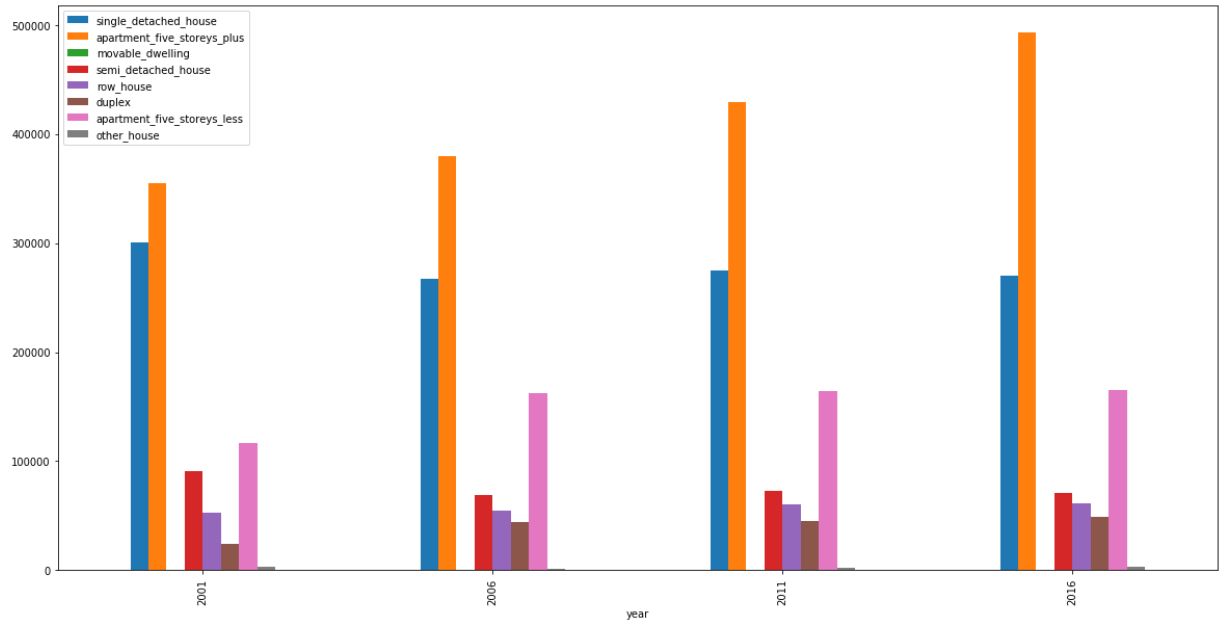


In [11]:
```
# Dwelling types - 2016 (Red)
create_bar_chart('2001', dwelling_types.iloc[3],'','', 'orange')
```



In [12]:
```
# Bar Plot
dwelling_types.plot.bar(figsize=(20,10))
```

Out[12]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1ded73f4ec8>
```



Average Monthly Shelter Costs in Toronto Per Year

In this section, you will calculate the average monthly shelter costs for owned and rented dwellings and the average house value for each year. Plot the results as a line chart.

Optional challenge: Plot each line chart in a different color.

In [13]:
```
# Calculate the average monthly shelter costs for owned and rented dwellings
shelter_costs = to_data.groupby(['year']).sum()
```

```
shelter_costs.head()
shelter_cost_owned = shelter_costs.iloc[:,-2]
shelter_cost_rented = shelter_costs.iloc[:,-1]
```

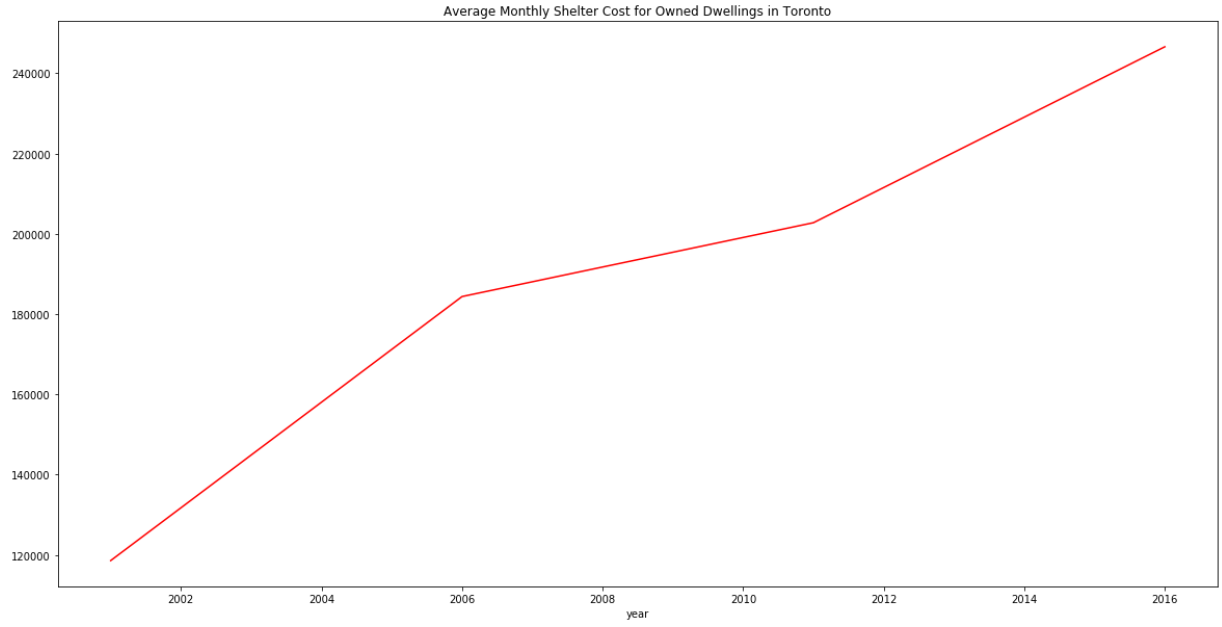
In [14]:

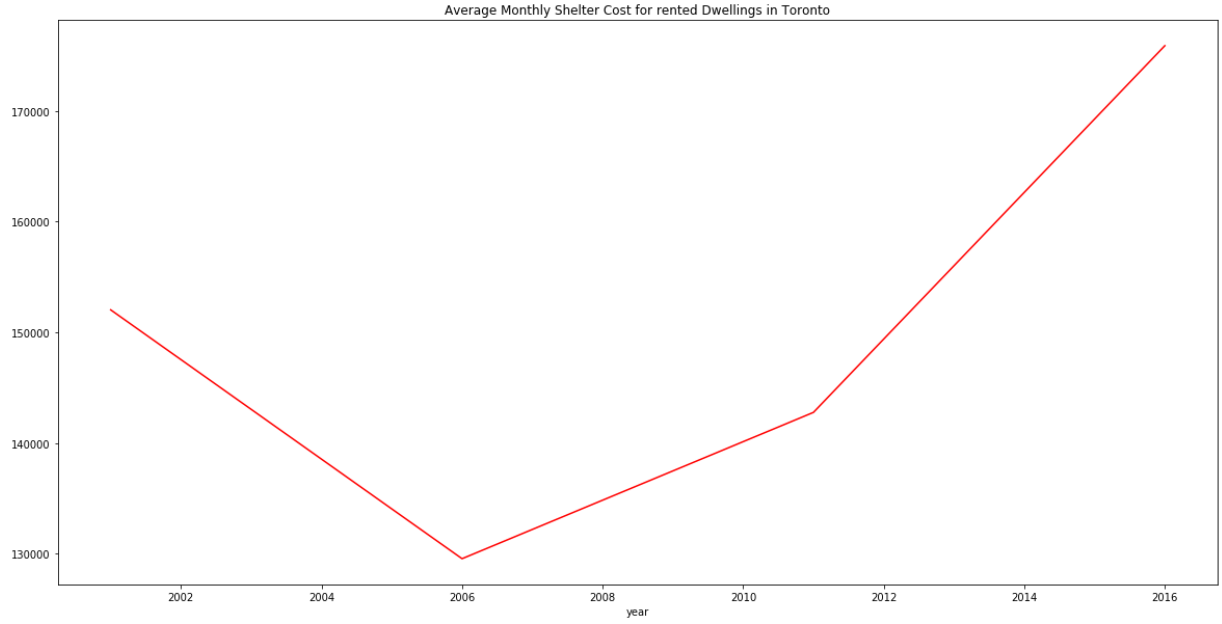
```
# Helper create_line_chart function
def create_line_chart(title, data, xlabel, ylabel, color):
    fig = plt.figure(figsize=(20,10))
    data.plot.line(title = title, color = color)
    plt.show()
```

In [15]:

```
# Create two line charts, one to plot the monthly shelter costs for owned dwelleing
# Line chart for owned dwellings
create_line_chart('Average Monthly Shelter Cost for Owned Dwellings in Toronto', she

# Line chart for rented dwellings
create_line_chart('Average Monthly Shelter Cost for rented Dwellings in Toronto', sh
```





Average House Value per Year

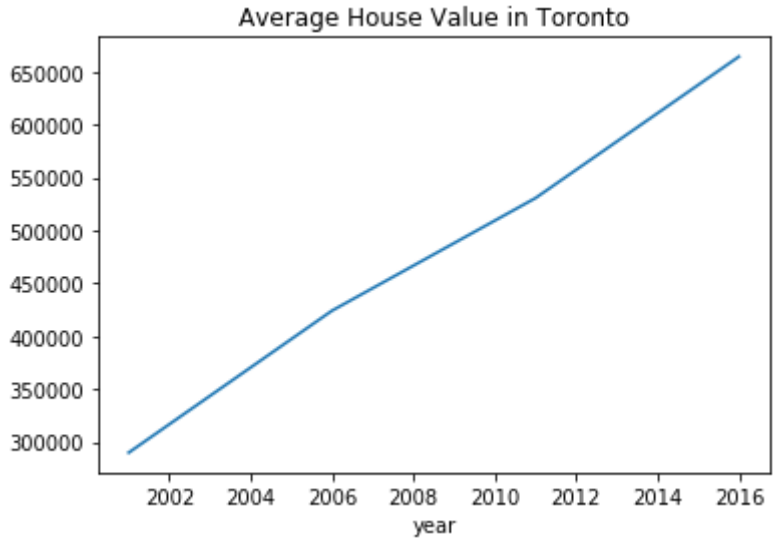
In this section, you want to determine the average house value per year. An investor may want to understand better the sales price of the rental property over time. For example, a customer will want to know if they should expect an increase or decrease in the property value over time so they can determine how long to hold the rental property. You will visualize the `average_house_value` per year as a bar chart.

In [16]:
```
# Calculate the average house value per year
housing_units = to_data.groupby(['year']).mean()
avg_house_value = housing_units.iloc[:,-3]
avg_house_value.head()
```

Out[16]:
```
year
2001    289882.885714
2006    424059.664286
2011    530424.721429
2016    664068.328571
Name: average_house_value, dtype: float64
```

In [17]:
```
# Plot the average house value per year as a line chart
avg_house_value.plot.line(title = 'Average House Value in Toronto')
```

Out[17]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1ded4f00348>
```



Average House Value by Neighbourhood

In this section, you will use `hvplot` to create an interactive visualization of the average house value with a dropdown selector for the neighbourhood.

Hint: It will be easier to create a new DataFrame from grouping the data and calculating the mean house values for each year and neighbourhood.

In [18]:
```
# Create a new DataFrame with the mean house values by neighbourhood per year
to_new_data = to_data.groupby([to_data.index, "neighbourhood"]).mean()
to_house_value = to_new_data["average_house_value"]
to_house_value = pd.DataFrame(to_house_value).reset_index()
to_house_value.head()
```

Out[18]:

	year	neighbourhood	average_house_value
0	2001	Agincourt North	200388.0
1	2001	Agincourt South-Malvern West	203047.0
2	2001	Alderwood	259998.0
3	2001	Annex	453850.0
4	2001	Banbury-Don Mills	371864.0

In [19]:

```
# Use hvplot to create an interactive line chart of the average house value per neig
to_house_value.hvplot(
    x='year',
    xlabel = 'Year',
    y='average_house_value',
    ylabel = 'Avg. House Value',
    groupby = 'neighbourhood',
    kind = 'line')
```

Out[19]:

Number of Dwelling Types per Year

In this section, you will use `hvplot` to create an interactive visualization of the average number of dwelling types per year with a dropdown selector for the neighbourhood.

In [20]:

```
# Fetch the data of all dwelling types per year
to_data.head()
dwelling_type = to_data.groupby([to_data.index, "neighbourhood"]).mean()

# Use hvplot function to display visualization of number of dwelling types in TO
dwelling_type.hvplot(
    x='year',
    xlabel = 'Year',
    y=['single_detached_house',
       'apartment_five_storeys_plus',
       'movable_dwelling',
       'semi_detached_house',
       'row_house',
       'duplex',
       'apartment_five_storeys_less',
       'other_house'],
    ylabel = 'Avg. House Value',
    groupby = 'neighbourhood',
    kind = 'bar',
    stacked = False,
    rot = 90,
    height = 500)
```

Out[20]:

In [21]:

```
# Use hvplot to create an interactive bar chart of the number of dwelling types per
dwelling_type.hvplot(
    x='year',
    xlabel = 'Year',
```

```
    y='average_house_value',
    ylabel = 'Avg. House Value',
    groupby = 'neighbourhood',
    kind = 'bar',
    stacked = False)
```

Out[21]:

The Top 10 Most Expensive Neighbourhoods

In this section, you will need to calculate the house value for each neighbourhood and then sort the values to obtain the top 10 most expensive neighbourhoods on average. Plot the results as a bar chart.

In [22]:
```
# Getting the data from the top 10 expensive neighbourhoods
top_10_most_expensive = to_data.sort_values(by='average_house_value', ascending=Fals
top_10_most_expensive
```

Out[22]:

year	neighbourhood	single_detached_house	apartment_five_storeys_plus	movable_dwelling	semi_d
2016	Bridle Path-Sunnybrook-York Mills	2275	590	0	
2011	Bridle Path-Sunnybrook-York Mills	2285	480	0	
2016	Forest Hill South	1685	2025	0	
2016	Lawrence Park South	3420	925	0	
2016	Rosedale-Moore Park	2450	4990	0	
2016	St.Andrew-Windfields	3245	1745	0	
2016	Casa Loma	875	2680	0	
2006	Bridle Path-Sunnybrook-York Mills	2205	145	0	
2011	Forest Hill South	1730	1825	0	
2016	Bedford Park-Nortown	4820	1995	0	

In [23]:
```
# Plotting the data from the top 10 expensive neighbourhoods
top_10_most_expensive.hvplot.bar(
    x="neighbourhood",
    xlabel = "Year",
    y="average_house_value",
    ylabel = "Avg. House Value",
```

```
    title = "Top 10 Most Expensive Neighborhoods in Toronto",
    height=500,
    rot=46)
```

Out[23]:

Neighbourhood Map

In this section, you will read in neighbourhoods location data and build an interactive map with the average house value per neighbourhood. Use a `scatter_mapbox` from Plotly express to create the visualization. Remember, you will need your Mapbox API key for this.

Load Location Data

In [24]:
```
# Load neighbourhoods coordinates data
file_path = Path("Data/toronto_neighbourhoods_coordinates.csv")
df_neighbourhood_locations = pd.read_csv(file_path)
df_neighbourhood_locations.head()
```

Out[24]:

	neighbourhood	lat	lon
0	Agincourt North	43.805441	-79.266712
1	Agincourt South-Malvern West	43.788658	-79.265612
2	Alderwood	43.604937	-79.541611
3	Annex	43.671585	-79.404001
4	Banbury-Don Mills	43.737657	-79.349718

Data Preparation

You will need to join the location data with the mean values per neighbourhood.

1. Calculate the mean values for each neighbourhood.

2. Join the average values with the neighbourhood locations.

In [25]:
```
# Calculate the mean values for each neighborhood
all_neighbourhoods = to_data.groupby("neighbourhood").mean()
all_neighbourhoods.head()
```

Out[25]:

neighbourhood	single_detached_house	apartment_five_storeys_plus	movable_dwelling	semi_detached
Agincourt North	3435.00	1947.50	2.50	
Agincourt South-Malvern West	2897.50	2180.00	1.25	
Alderwood	2903.75	302.50	1.25	

neighbourhood	single_detached_house	apartment_five_storeys_plus	movable_dwelling	semi_detached
Annex	751.25	7235.00	1.25	
Banbury-Don Mills	3572.50	5388.75	1.25	

In [26]:
```
# Joint the average values with the combined neighbourhood locations
combined_df = pd.merge(df_neighbourhood_locations, all_neighbourhoods, on="neighbour
combined_df.head()
```

Out[26]:

	neighbourhood	lat	lon	single_detached_house	apartment_five_storeys_plus	movab
0	Agincourt North	43.805441	-79.266712	3435.00	1947.50	
1	Agincourt South-Malvern West	43.788658	-79.265612	2897.50	2180.00	
2	Alderwood	43.604937	-79.541611	2903.75	302.50	
3	Annex	43.671585	-79.404001	751.25	7235.00	
4	Banbury-Don Mills	43.737657	-79.349718	3572.50	5388.75	

Mapbox Visualization

Plot the average values per neighbourhood using a Plotly express `scatter_mapbox` visualization.

In [27]:
```
# Create a scatter mapbox to analyze neighbourhood info
map_plot = px.scatter_mapbox(
    combined_df,
    lat="lat",
    lon="lon",
    size="average_house_value",
    hover_name="neighbourhood",
)

map_plot.show()
```

In [28]:
```
combined_df.head()
```

Out[28]:

	neighbourhood	lat	lon	single_detached_house	apartment_five_storeys_plus	movab
0	Agincourt North	43.805441	-79.266712	3435.00	1947.50	
1	Agincourt South-Malvern West	43.788658	-79.265612	2897.50	2180.00	
2	Alderwood	43.604937	-79.541611	2903.75	302.50	
3	Annex	43.671585	-79.404001	751.25	7235.00	
4	Banbury-Don Mills	43.737657	-79.349718	3572.50	5388.75	

Cost Analysis - Optional Challenge

In this section, you will use Plotly express to a couple of plots that investors can interactively filter and explore various factors related to the house value of the Toronto's neighbourhoods.

Create a bar chart row facet to plot the average house values for all Toronto's neighbourhoods per year

In [29]:
```
# Create bar chart row facet plot
to_data.reset_index(level=0, inplace=True)
fig = px.bar(
    to_data,
    x="neighbourhood",
    facet_row_spacing=0.08,
    facet_col_spacing=1.00,
    height=1100, width=1300,
    y=['single_detached_house',
       'apartment_five_storeys_plus',
       'movable_dwelling',
       'semi_detached_house',
       'row_house',
```

```
        'duplex',
        'apartment_five_storeys_less',
        'other_house'
      ],

    color="average_house_value",
    facet_row = 'year',
)
fig.show()
```

Create a sunburst chart to conduct a costs analysis of most expensive neighbourhoods in Toronto per year

In [30]:
```
# Getting the data from the top 10 expensive neighbourhoods
most_expensive_homes = to_data.sort_values(by='average_house_value', ascending=False
most_expensive_homes.head()
```

Out[30]:

	year	neighbourhood	single_detached_house	apartment_five_storeys_plus	movable_dwelling	se
436	2016	Bridle Path-Sunnybrook-York Mills	2275	590	0	
296	2011	Bridle Path-Sunnybrook-York Mills	2285	480	0	
464	2016	Forest Hill South	1685	2025	0	
489	2016	Lawrence Park South	3420	925	0	
524	2016	Rosedale-Moore Park	2450	4990	0	

In [31]:
```
# Fetch the data from all expensive neighbourhoods per year.
most_expensive_homes = to_data.drop([
    'single_detached_house',
    'apartment_five_storeys_plus',
    'movable_dwelling',
    'semi_detached_house',
    'row_house', 'duplex',
    'apartment_five_storeys_less',
    'other_house',
    'average_house_value'], axis = 1)
```

In [32]:
```
# Inspect DataFrame
most_expensive_homes.head()
```

Out[32]:

	year	neighbourhood	shelter_costs_owned	shelter_costs_rented
0	2001	Agincourt North	810	870

	year	neighbourhood	shelter_costs_owned	shelter_costs_rented
1	2001	Agincourt South-Malvern West	806	892
2	2001	Alderwood	817	924
3	2001	Annex	1027	1378
4	2001	Banbury-Don Mills	1007	1163

In [33]:

```
# Display sunburst plot
fig = px.sunburst(
    most_expensive_homes,
    path = ['year','neighbourhood'],
    values = 'shelter_costs_owned',
    color = 'shelter_costs_owned',
    hover_data = ['shelter_costs_owned', 'shelter_costs_rented'],
    color_continuous_scale = 'sunset',
    height = 500,
    title = "Most Expensive Homes in Toronto, Canada"
)
fig.show()
```