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]:
```
# Imported the necesscary libraries for project

import panel as pn
pn.extension('plotly')

import plotly.express as px

import plotly

import pandas as pd

import hvplot.pandas

from matplotlib import pyplot as plt

import os

from pathlib import Path

from dotenv import load_dotenv

import numpy as np
```

In [2]:
```
# Loaded .env environment variables

load_dotenv()

map_box_api = os.getenv("mapbox_akey")

px.set_mapbox_access_token(map_box_api)
```

Load Data

In [3]:
```
# Loaded the census data into a Pandas DataFrame

file_path = Path("Data/toronto_neighbourhoods_census_data.csv")

census_data = pd.read_csv(file_path, index_col="year")

census_data.head()
```

Out[3]:

	neighbourhood	single_detached_house	apartment_five_storeys_plus	movable_dwelling	semi

year	neighbourhood	single_detached_house	apartment_five_storeys_plus	movable_dwelling	semi
year					
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]:
```
# Calculated the sum number of dwelling types units per year using groupby functio

dwelling_df = census_data[["neighbourhood", "single_detached_house", "apartment_fi

dwelling_df
```

Out[4]:

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

In [5]:
```
# Saved the dataframe as a csv file

dwelling_df.to_csv("dwelling.csv")
```

In [6]:
```
# Extracted data for each from "dwelling_df"

year_2001 = pd.DataFrame(dwelling_df.iloc[0,:])
```

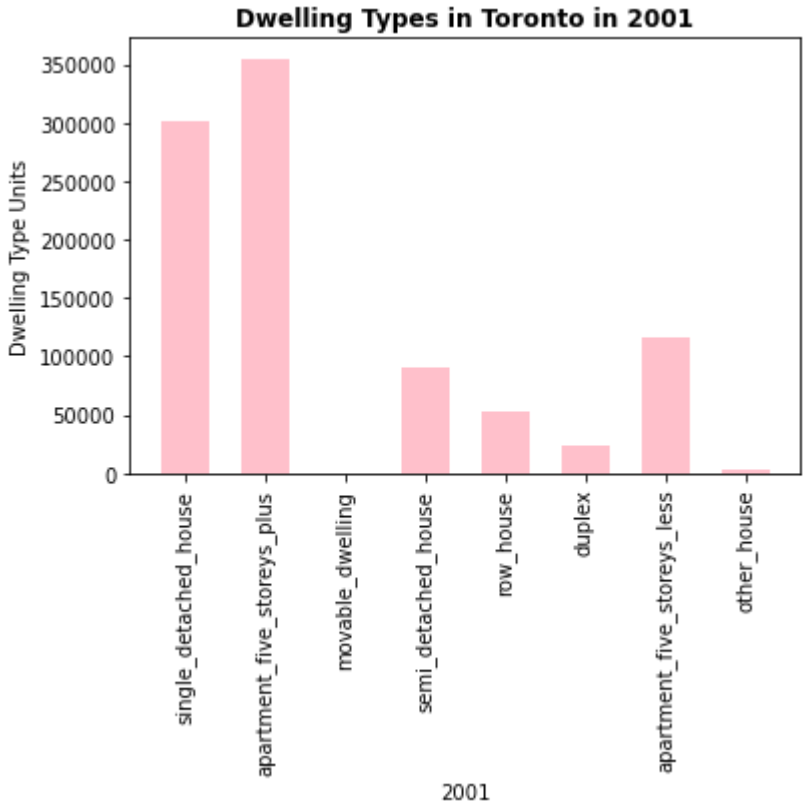
```
year_2001.rename(columns={2001:'dwelling_number'}, inplace=True)

year_2006 = pd.DataFrame(dwelling_df.iloc[1,:])

year_2006.rename(columns={2006:'dwelling_number'}, inplace=True)

year_2011 = pd.DataFrame(dwelling_df.iloc[2,:])

year_2011.rename(columns={2011:'dwelling_number'}, inplace=True)

year_2016 = pd.DataFrame(dwelling_df.iloc[3,:])

year_2016.rename(columns={2016:'dwelling_number'}, inplace=True)
```

In [7]:

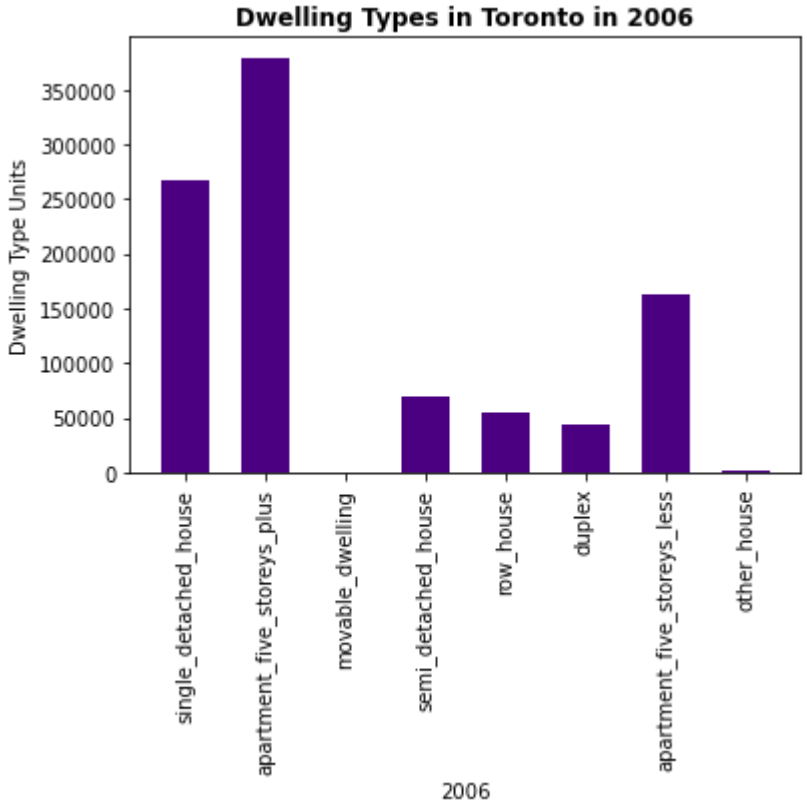
```
# Created bar_chart function

def bar_chart(data, title, xlabel, ylabel, color):
    plt.bar(x=data.index, height=data.dwelling_number, color=color, width = 0.6)
    plt.title(title,fontweight='bold')
    plt.xlabel(xlabel)
    plt.ylabel(ylabel)
    plt.xticks(rotation=90)
```

In [8]:

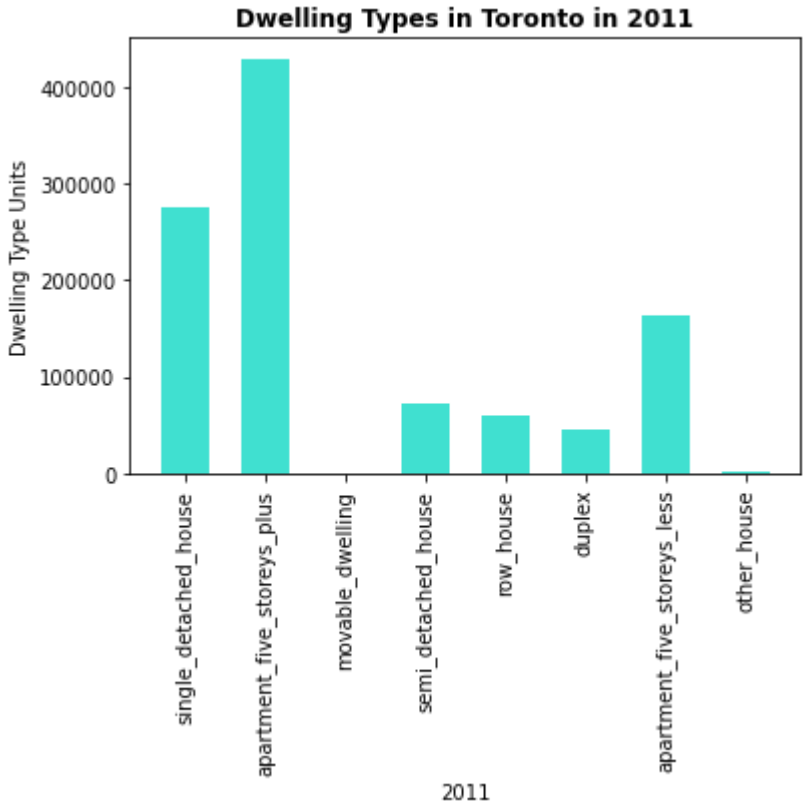
```
# Createed a bar chart per year to show the number of dwelling types

# Bar chart for 2001

bar_chart(year_2001, 'Dwelling Types in Toronto in 2001', '2001', 'Dwelling Type U
```



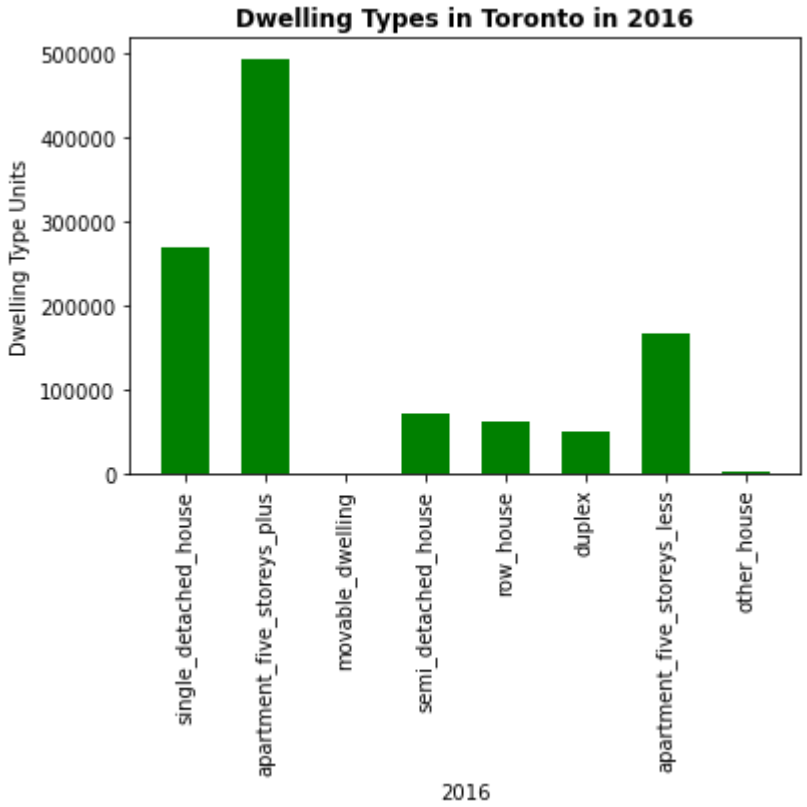
In [9]:
```
# Bar chart for 2006

bar_chart(year_2006, 'Dwelling Types in Toronto in 2006', '2006', 'Dwelling Type U
```



In [10]:
```
# Bar chart for 2011

bar_chart(year_2011, 'Dwelling Types in Toronto in 2011', '2011', 'Dwelling Type U
```



In [11]:
```
# Bar chart for 2016

bar_chart(year_2016, 'Dwelling Types in Toronto in 2016', '2016', 'Dwelling Type U
```



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.

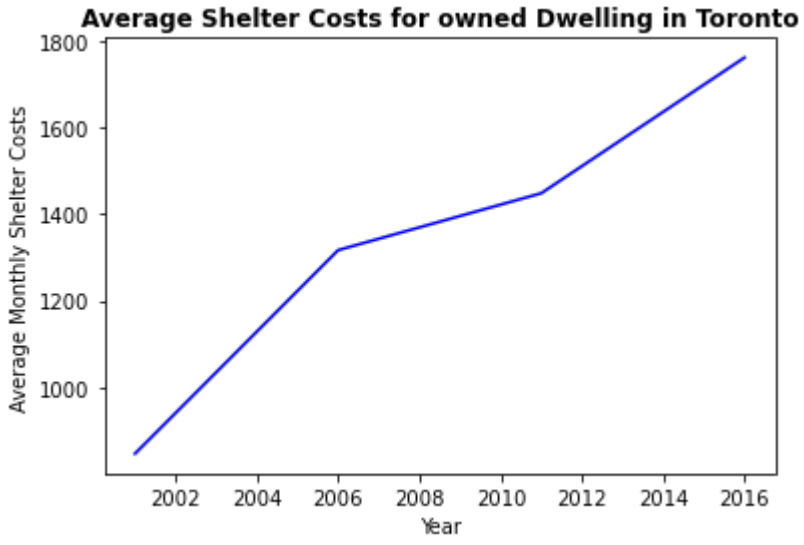
Optional challenge: Plot each line chart in a different color.

In [12]:
```
# Calculated the average monthly shelter costs for owned and rented dwellings

shelter_avg_costs = census_data[['shelter_costs_owned', 'shelter_costs_rented']].g

shelter_avg_costs
```

Out[12]:

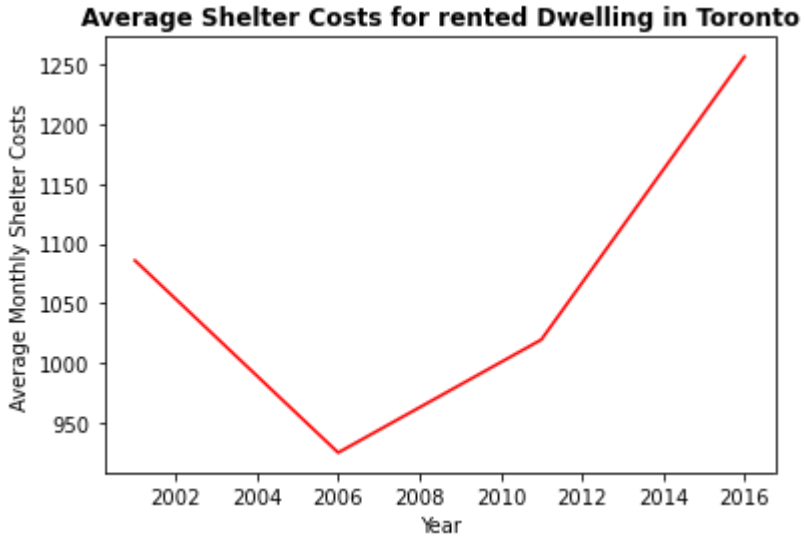
year	shelter_costs_owned	shelter_costs_rented
2001	846.878571	1085.935714
2006	1316.800000	925.414286
2011	1448.214286	1019.792857
2016	1761.314286	1256.321429

In [13]:
```
# Created line_chart function

def line_chart(data, title, xlabel, ylabel, color):
    plt.plot(data, color=color)
    plt.xlabel(xlabel)
    plt.ylabel(ylabel)
    plt.title(title, fontweight='bold')

# Plotted line chart for owned dwellings

line_chart(shelter_avg_costs['shelter_costs_owned'],'Average Shelter Costs for own
```



In [14]:
```

```

```
# Plotted line chart for rented dwellings

line_chart(shelter_avg_costs['shelter_costs_rented'], 'Average Shelter Costs for r
```



Average House Value per Year

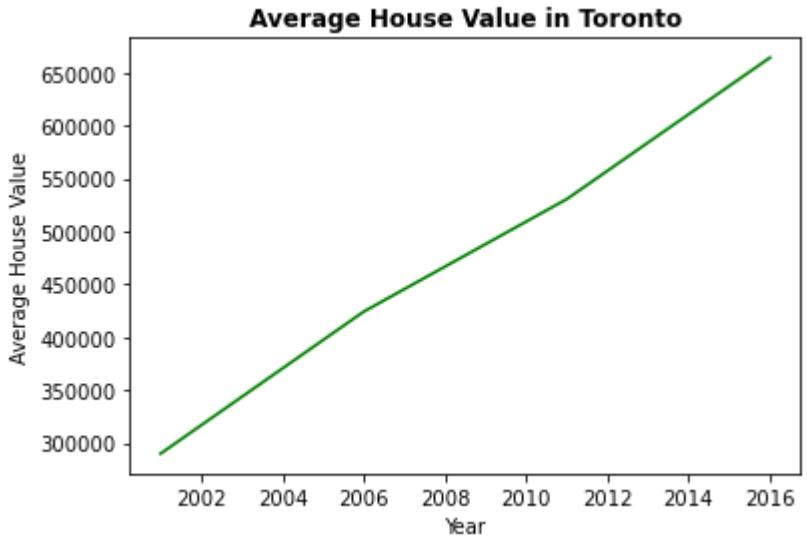
In this section, you want to determine the average house value per year. An investor may want to understand better the sales price of the rental property over time. For example, a customer will want to know if they should expect an increase or decrease in the property value over time so they can determine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

In [15]:
```
# Calculate the average house value per year

avg_hour_value = census_data[['average_house_value']].groupby(census_data.index).m

avg_hour_value
```

Out[15]:

year	average_house_value
2001	289882.885714
2006	424059.664286
2011	530424.721429
2016	664068.328571

In [16]:
```
# Plotted the average house value per year as a line chart

line_chart(avg_hour_value, 'Average House Value in Toronto', 'Year', 'Average Hous
```



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.

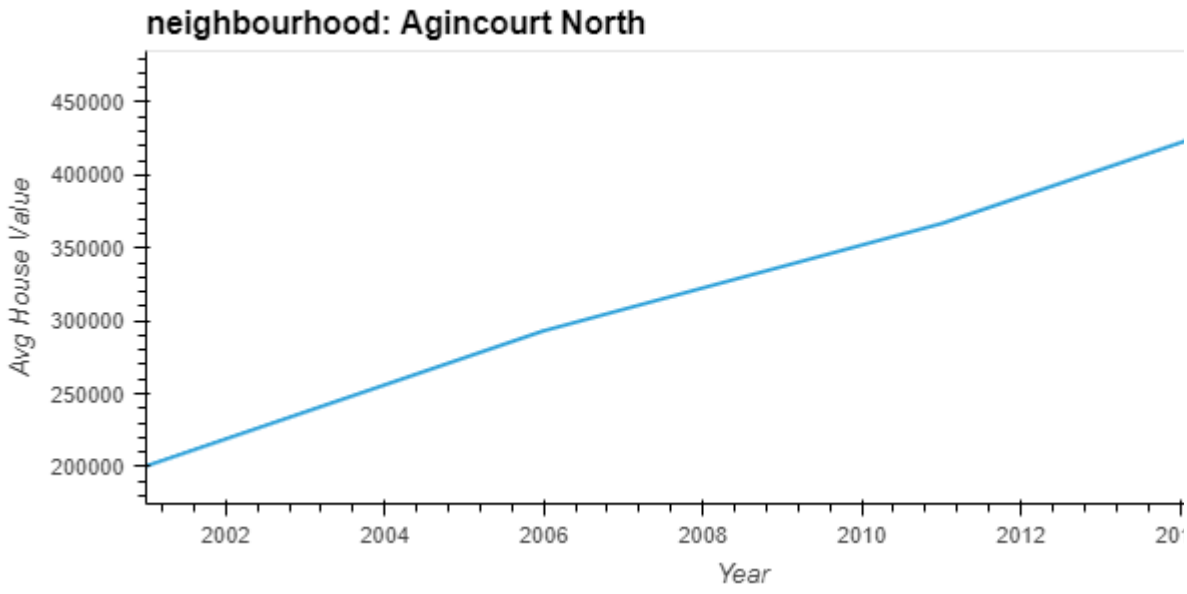
In [17]:
```
# Created a new DataFrame with the mean house values by neighbourhood per year

avg_by_neighbourhood = census_data[['neighbourhood','average_house_value']].groupb

avg_by_neighbourhood.set_index('year', inplace=True)

avg_by_neighbourhood[:10]
```

Out[17]:

year	neighbourhood	average_house_value
2001	Agincourt North	200388
2001	Agincourt South-Malvern West	203047
2001	Alderwood	259998
2001	Annex	453850
2001	Banbury-Don Mills	371864
2001	Bathurst Manor	304749
2001	Bay Street Corridor	257404
2001	Bayview Village	327644
2001	Bayview Woods-Steeles	343535
2001	Bedford Park-Nortown	565304

In [18]:
```
# Created an interactive line chart of the average house value per neighbourhood u

avg_by_neighbourhood.hvplot(groupby='neighbourhood', xlabel='Year', ylabel='Avg Ho
```

Out[18]:



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.

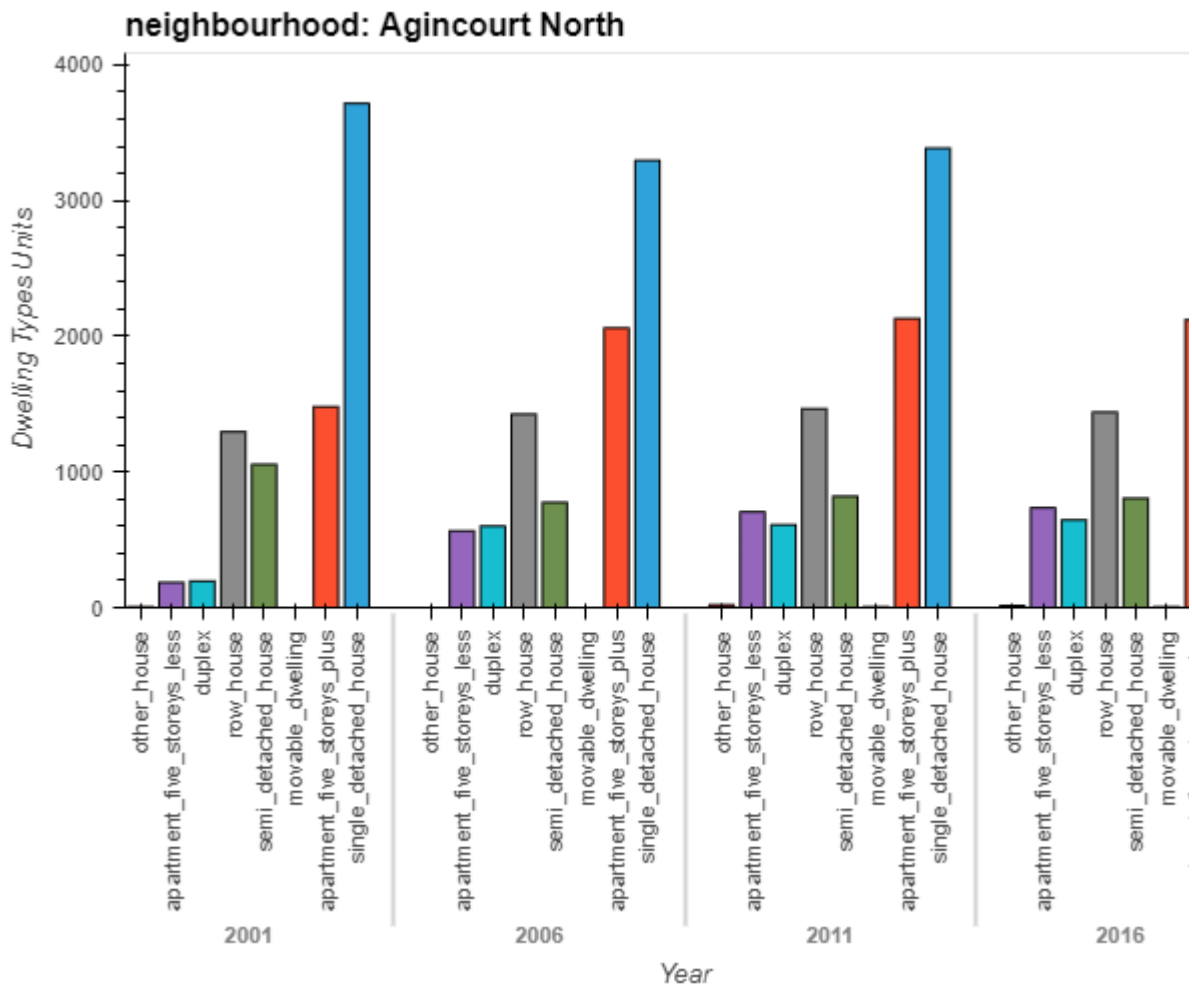
In [19]:
```
# Fetched the data of all dwelling types per year

number_dwelling_per_year = census_data[["neighbourhood",'single_detached_house','a

number_dwelling_per_year.set_index("year", inplace=True)

number_dwelling_per_year[:5]
```

Out[19]:

year	neighbourhood	single_detached_house	apartment_five_storeys_plus	movable_dwelling	semi
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	

In [20]:
```
# Created an interactive line chart of the number of dwelling types per neighbourh

number_dwelling_per_year.hvplot.bar(groupby="neighbourhood", rot=90,width=650, hei
```

Out[20]:



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 [21]:
```
# Fetched data from the top 10 expensive neighbourhoods

expensive_house = census_data[["neighbourhood",'single_detached_house','apartment_

expensive_house = expensive_house[:10]

# Initialized Data

expensive_house.reset_index(inplace=True)

expensive_house
```

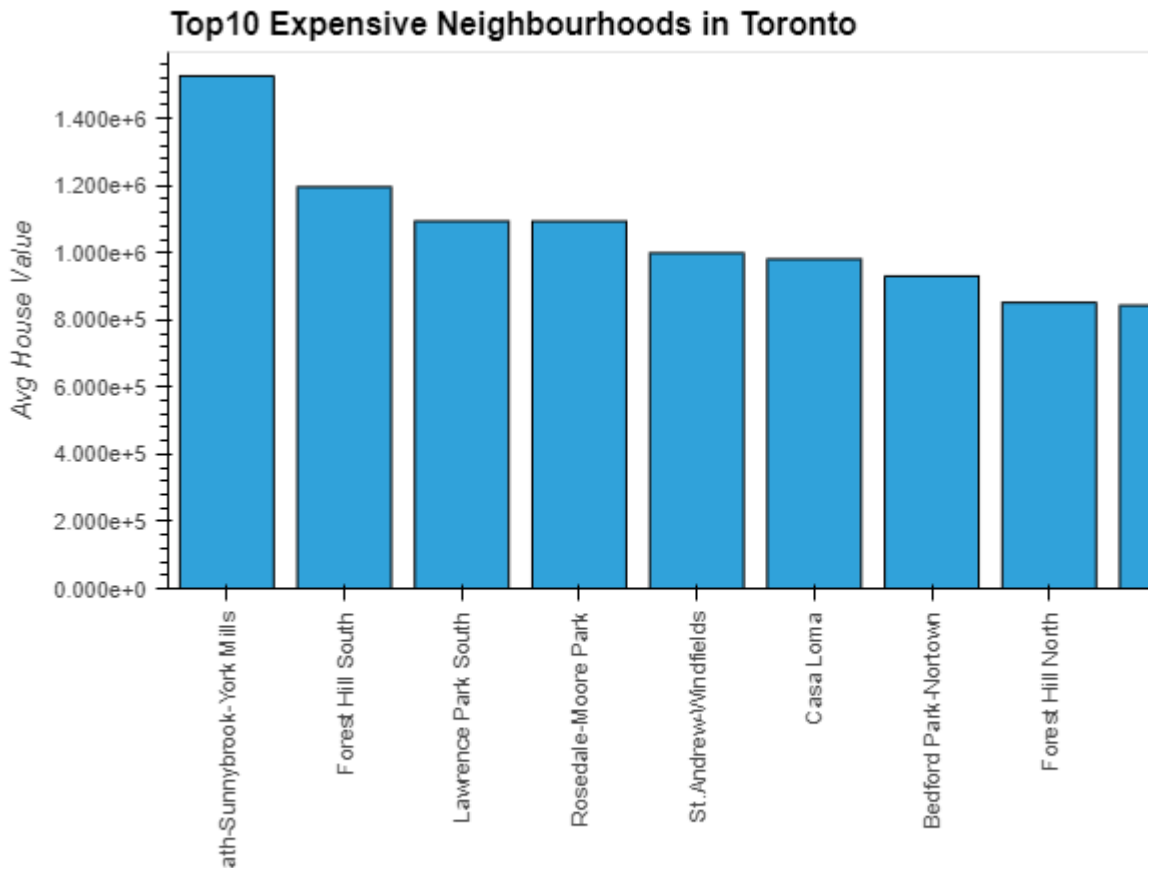
Out[21]:

	neighbourhood	single_detached_house	apartment_five_storeys_plus	movable_dwelling	semi_det
0	Bridle Path-Sunnybrook-York Mills	2260.00	331.25	0.00	
1	Forest Hill South	1742.50	2031.25	1.25	
2	Lawrence Park South	3472.50	773.75	0.00	
3	Rosedale-Moore Park	2498.75	4641.25	0.00	
4	St.Andrew-Windfields	3225.00	1670.00	0.00	
5	Casa Loma	916.25	2310.00	0.00	
6	Bedford Park-Nortown	4865.00	1981.25	0.00	
7	Forest Hill North	1488.75	3392.50	0.00	
8	Kingsway South	2326.25	576.25	0.00	
9	Yonge-St.Clair	565.00	3948.75	0.00	

In [22]:

```
# Plotted the data from the top 10 expensive neighbourhoods

expensive_house.hvplot.bar(x='neighbourhood', y='average_house_value', rot=90, hei
```

Out[22]:



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 [23]:
```
# 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[23]:

	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 [24]:
```
# Calculated the mean values for each neighborhood

mean_neighborhood =  census_data[['neighbourhood','single_detached_house','apartme

mean_neighborhood.head()
```

Out[24]:

	neighbourhood	single_detached_house	apartment_five_storeys_plus	movable_dwelling	semi_det
0	Agincourt North	3435.00	1947.50	2.50	

	neighbourhood	single_detached_house	apartment_five_storeys_plus	movable_dwelling	semi_det
1	Agincourt South-Malvern West	2897.50	2180.00	1.25	
2	Alderwood	2903.75	302.50	1.25	
3	Annex	751.25	7235.00	1.25	
4	Banbury-Don Mills	3572.50	5388.75	1.25	

In [25]:
```
# Joined the average values with the neighbourhood locations

combined_df = pd.concat([df_neighbourhood_locations.set_index('neighbourhood'),mea

combined_df.head()

combined_df.reset_index(inplace=True)
```

Mapbox Visualization

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

In [26]:
```
# Created a scatter mapbox to analyze neighbourhood info

map_1 = px.scatter_mapbox(
    combined_df,
    lat="lat",
    lon="lon",
    color="average_house_value",
    size="average_house_value",
    color_continuous_scale=px.colors.sequential.Cividis,
    zoom = 10
)

map_1.show()
```

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 [27]:
```
# Recalled and checked dataframe "avg_by_neighbourhood"
```

```
avg_by_neighbourhood.head()
```

Out[27]:

year	neighbourhood	average_house_value
2001	Agincourt North	200388
2001	Agincourt South-Malvern West	203047
2001	Alderwood	259998
2001	Annex	453850
2001	Banbury-Don Mills	371864

In [28]:

```
# Created a bar chart row facet to plot the average house values for all Toronto's

px.bar(avg_by_neighbourhood,x='neighbourhood',y='average_house_value',facet_row=av
```

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

In [29]:

```
# Fetched the data from all expensive neighbourhoods per year.

costs_per_neighbourhood = census_data.filter(["neighbourhood","shelter_costs_owned

costs_per_neighbourhood
```

Out[29]:

year	neighbourhood	shelter_costs_owned	shelter_costs_rented
2001	Agincourt North	810	870
2001	Agincourt South-Malvern West	806	892
2001	Alderwood	817	924
2001	Annex	1027	1378
2001	Banbury-Don Mills	1007	1163
...	...	...	...
2016	Wychwood	1864	1146
2016	Yonge-Eglinton	2398	1535
2016	Yonge-St.Clair	2192	1619
2016	York University Heights	1444	1122
2016	Yorkdale-Glen Park	1451	1128

560 rows × 3 columns

In [30]:

```
# Created a dataframe for top neighborhood

top_neighbourhood = expensive_house.reset_index()

top_neighbourhood = top_neighbourhood["neighbourhood"].tolist()

top_neighbourhood
```

Out[30]:
```
['Bridle Path-Sunnybrook-York Mills',
 'Forest Hill South',
 'Lawrence Park South',
 'Rosedale-Moore Park',
 'St.Andrew-Windfields',
 'Casa Loma',
 'Bedford Park-Nortown',
 'Forest Hill North',
 'Kingsway South',
 'Yonge-St.Clair']
```

In [31]:
```
# Initiated a query

top_neighbourhood_df = costs_per_neighbourhood.query("neighbourhood in @top_neighb

top_neighbourhood_df
```

Out[31]:

year	neighbourhood	shelter_costs_owned	shelter_costs_rented
2001	Bedford Park-Nortown	933	1491
2001	Bridle Path-Sunnybrook-York Mills	1983	1790
2001	Casa Loma	1241	1500
2001	Forest Hill North	940	1428
2001	Forest Hill South	1001	1469
2001	Kingsway South	1362	1340
2001	Lawrence Park South	1021	1630
2001	Rosedale-Moore Park	1219	1540
2001	St.Andrew-Windfields	1055	1551
2001	Yonge-St.Clair	1091	1351
2006	Bedford Park-Nortown	1808	1020
2006	Bridle Path-Sunnybrook-York Mills	2170	2167
2006	Casa Loma	1819	1356
2006	Forest Hill North	1731	1027
2006	Forest Hill South	1781	1094
2006	Kingsway South	1625	1488
2006	Lawrence Park South	1977	1116
2006	Rosedale-Moore Park	1868	1332

year	neighbourhood	shelter_costs_owned	shelter_costs_rented
2006	St.Andrew-Windfields	1881	1152
2006	Yonge-St.Clair	1638	1192
2011	Bedford Park-Nortown	1988	1124
2011	Bridle Path-Sunnybrook-York Mills	2387	2388
2011	Casa Loma	2001	1494
2011	Forest Hill North	1904	1132
2011	Forest Hill South	1959	1206
2011	Kingsway South	1787	1640
2011	Lawrence Park South	2174	1230
2011	Rosedale-Moore Park	2054	1468
2011	St.Andrew-Windfields	2069	1270
2011	Yonge-St.Clair	1802	1314
2016	Bedford Park-Nortown	2418	1385
2016	Bridle Path-Sunnybrook-York Mills	2903	2942
2016	Casa Loma	2434	1841
2016	Forest Hill North	2316	1395
2016	Forest Hill South	2383	1486
2016	Kingsway South	2173	2020
2016	Lawrence Park South	2644	1515
2016	Rosedale-Moore Park	2498	1809
2016	St.Andrew-Windfields	2516	1565
2016	Yonge-St.Clair	2192	1619

In [32]:
```
# Prepared a new dataframe for sunburst chart called "df1"

df1 = top_neighbourhood_df.reset_index()

df1
```

Out[32]:

	year	neighbourhood	shelter_costs_owned	shelter_costs_rented
0	2001	Bedford Park-Nortown	933	1491
1	2001	Bridle Path-Sunnybrook-York Mills	1983	1790
2	2001	Casa Loma	1241	1500
3	2001	Forest Hill North	940	1428
4	2001	Forest Hill South	1001	1469

	year	neighbourhood	shelter_costs_owned	shelter_costs_rented
5	2001	Kingsway South	1362	1340
6	2001	Lawrence Park South	1021	1630
7	2001	Rosedale-Moore Park	1219	1540
8	2001	St.Andrew-Windfields	1055	1551
9	2001	Yonge-St.Clair	1091	1351
10	2006	Bedford Park-Nortown	1808	1020
11	2006	Bridle Path-Sunnybrook-York Mills	2170	2167
12	2006	Casa Loma	1819	1356
13	2006	Forest Hill North	1731	1027
14	2006	Forest Hill South	1781	1094
15	2006	Kingsway South	1625	1488
16	2006	Lawrence Park South	1977	1116
17	2006	Rosedale-Moore Park	1868	1332
18	2006	St.Andrew-Windfields	1881	1152
19	2006	Yonge-St.Clair	1638	1192
20	2011	Bedford Park-Nortown	1988	1124
21	2011	Bridle Path-Sunnybrook-York Mills	2387	2388
22	2011	Casa Loma	2001	1494
23	2011	Forest Hill North	1904	1132
24	2011	Forest Hill South	1959	1206
25	2011	Kingsway South	1787	1640
26	2011	Lawrence Park South	2174	1230
27	2011	Rosedale-Moore Park	2054	1468
28	2011	St.Andrew-Windfields	2069	1270
29	2011	Yonge-St.Clair	1802	1314
30	2016	Bedford Park-Nortown	2418	1385
31	2016	Bridle Path-Sunnybrook-York Mills	2903	2942
32	2016	Casa Loma	2434	1841
33	2016	Forest Hill North	2316	1395
34	2016	Forest Hill South	2383	1486
35	2016	Kingsway South	2173	2020
36	2016	Lawrence Park South	2644	1515
37	2016	Rosedale-Moore Park	2498	1809
38	2016	St.Andrew-Windfields	2516	1565

	year	neighbourhood	shelter_costs_owned	shelter_costs_rented
39	2016	Yonge-St.Clair	2192	1619

In [33]:
```
# Stored values of the df1 in seperate variables

year = df1['year']

neighbourhood = df1['neighbourhood']

shelter_costs_owned = df1['shelter_costs_owned']

shelter_costs_rented = df1['shelter_costs_rented']
```

In [34]:
```
# Created a chart and store variables inside fig

fig = px.sunburst(df1,
                  path=[year, neighbourhood],
                  values=shelter_costs_owned,
                  color=shelter_costs_owned,
                  hover_data=[shelter_costs_rented],
                  color_continuous_scale=['red', 'yellow', 'green'],
                  title='Costs Analysis of Most Expensive Neighbourhood in Toronto pe
)
```

In [35]:
```
# Initialized sunburst

fig.show()
```