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 panel as pn
pn.extension('plotly')
import plotly.express as px
import pandas as pd
import hvplot.pandas
import matplotlib.pyplot as plt
import os
from pathlib import Path
from dotenv import load_dotenv
import random
```

In [2]:
```
# Read the Mapbox API key
load_dotenv()
map_box_api = os.getenv("MAPBOX_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_detached
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)
#We will also drop the columns which are not the dwelling types
dwellings_per_year = to_data.groupby("year").sum().drop(columns=["average_house_value", "s

dwellings_per_year
```

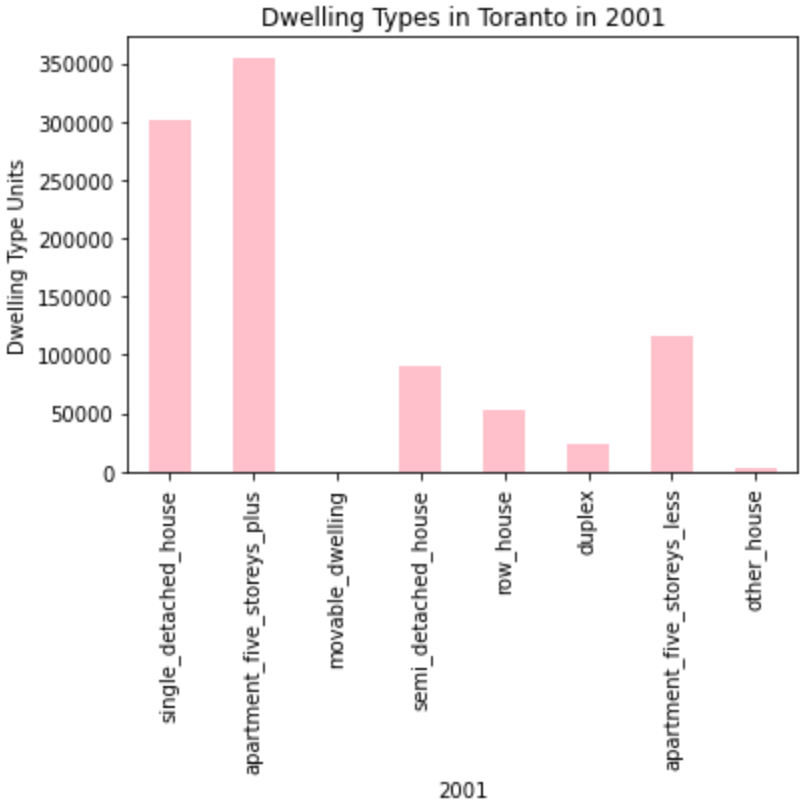
Out[4]:

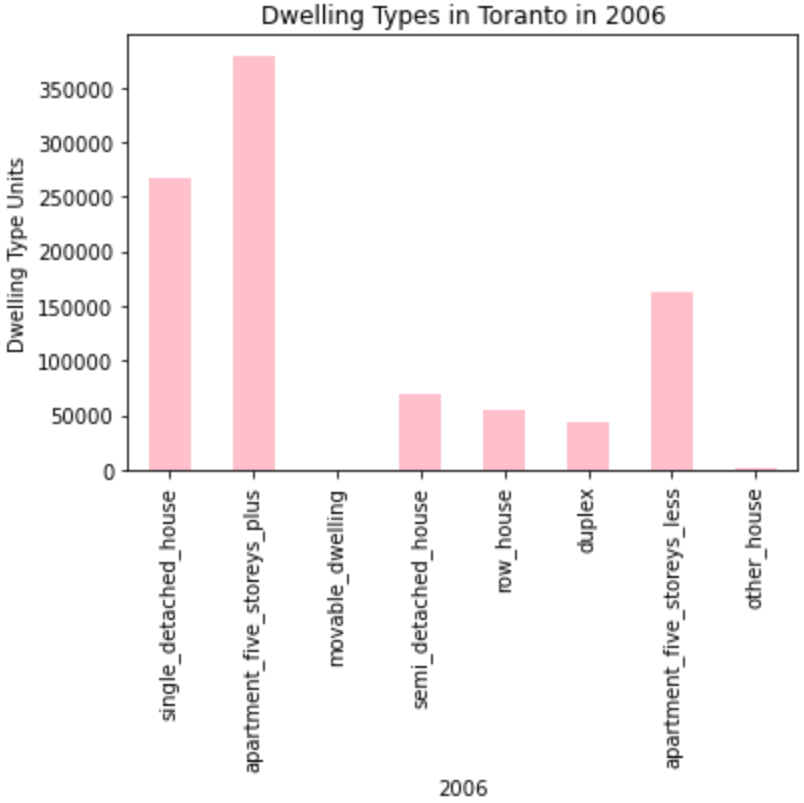
year	single_detached_house	apartment_five_storeys_plus	movable_dwelling	semi_detached_house	row_ho
2001	300930	355015	75	90995	52
2006	266860	379400	165	69430	54
2011	274940	429220	100	72480	60
2016	269680	493270	95	71200	61

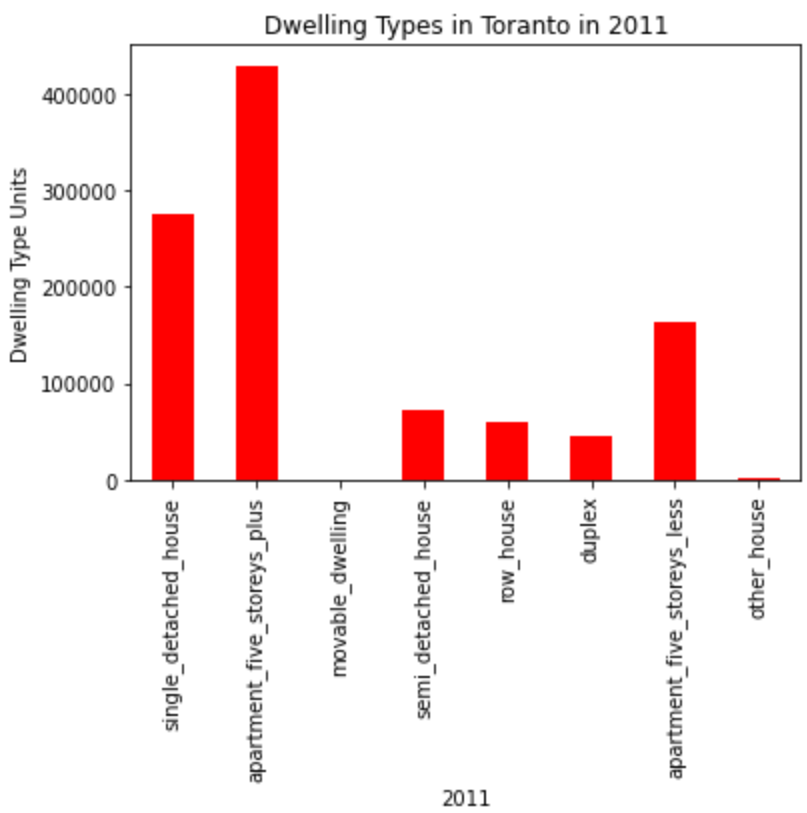
In [5]:
```
# Save the dataframe as a csv file
dwellings_per_year.to_csv("Data/dwelling_types_per_year.csv")
```

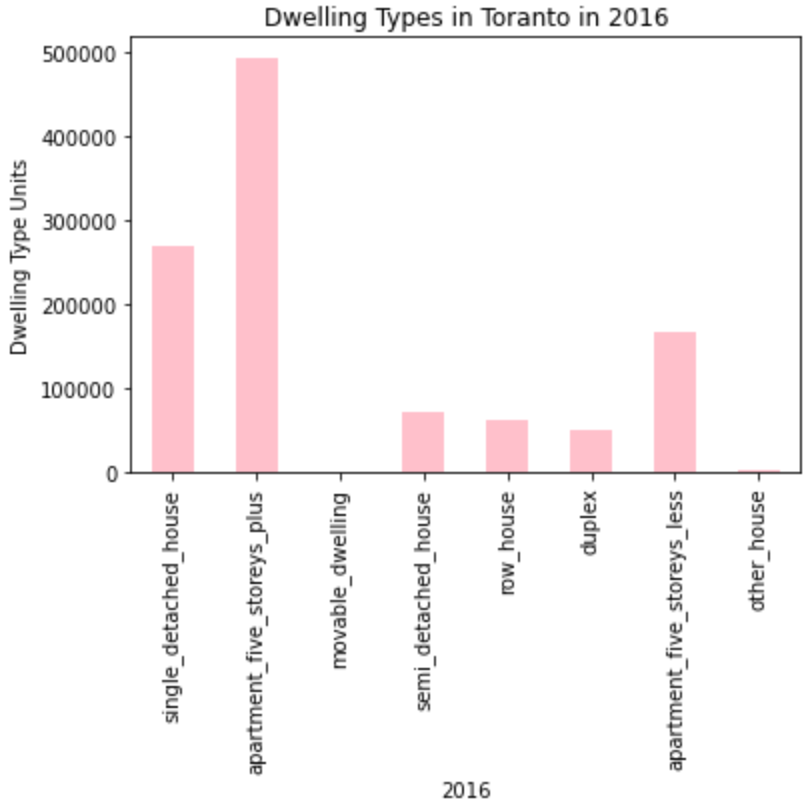
In [6]:
```
# Helper create_bar_chart function
def create_bar_chart(data, title, xlabel, ylabel, color):
    """
    Create a barplot based in the data argument.
    Input:
    data = DataFrame to use for plotting the data
    title = Chart Title
    xlabel = Label for X Axis
    ylabel = Label for Y Axis
    color = Colour of the bar chart

    """
    data.plot(kind="bar", xlabel=xlabel, ylabel=ylabel, color=color, title=title)

    #We want to show the plot after plotting it, if we dont do this, it will overwrite the
    plt.show()
```

In [7]:
```
# Create a bar chart per year to show the number of dwelling types
colors = ["red", "pink", "blue", "yellow"]

#We will use a for loop to do charts for each year rather than repeating the code and we w
for index in dwellings_per_year.index:
    title = "Dwelling Types in Toranto in " + str(index)
    ylabel = "Dwelling Type Units"
    create_bar_chart(dwellings_per_year.loc[index], title=title, xlabel=index, ylabel=ylab
```


Dwelling Types in Toranto in 2001
Dwelling Type Units
0
50000
100000
150000
200000
250000
300000
350000
single_detached_house
apartment_five_storeys_plus
movable_dwelling
semi_detached_house
row_house
duplex
apartment_five_storeys_less
other_house
2001


Dwelling Types in Toranto in 2006
Dwelling Type Units
0
50000
100000
150000
200000
250000
300000
350000
single_detached_house
apartment_five_storeys_plus
movable_dwelling
semi_detached_house
row_house
duplex
apartment_five_storeys_less
other_house
2006





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 [8]:

```
# Calculate the average monthly shelter costs for owned and rented dwellings
#We will also drop the columns which are not the dwelling types
```

```
#dwellings_per_year = to_data[[]].groupby("year").mean().drop(columns=["average_house_valu

monthly_shelter_cost_per_year = to_data[["shelter_costs_owned","shelter_costs_rented"]].gr
monthly_shelter_cost_per_year
```

Out[8]:

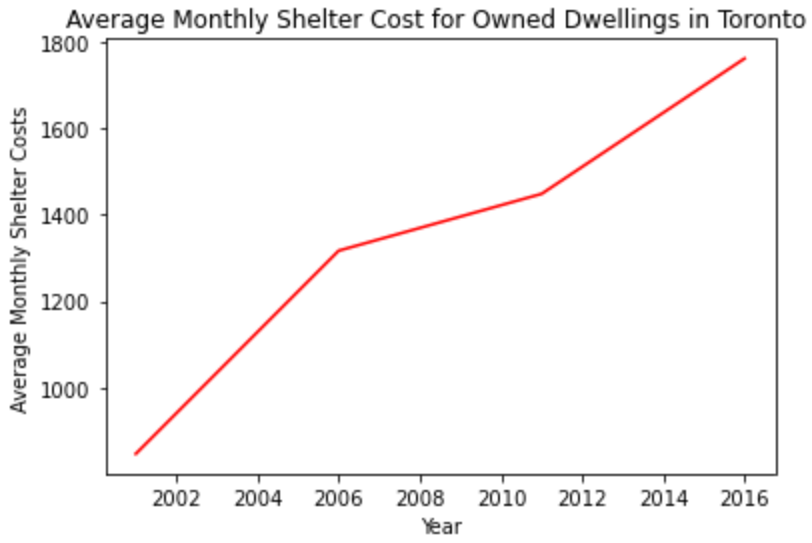
year	shelter_costs_owned	shelter_costs_rented
2001	846.878571	1085.935714
2006	1316.800000	925.414286
2011	1448.214286	1019.792857
2016	1761.314286	1256.321429

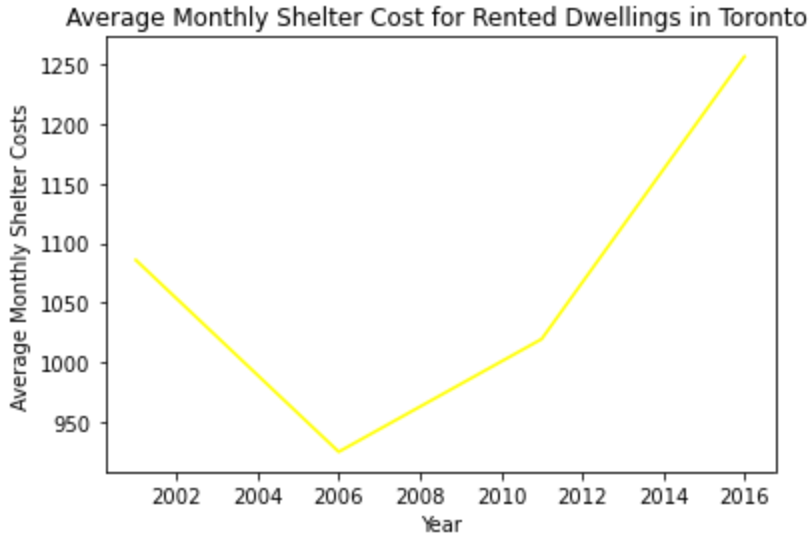
In [9]:

```
# Helper create_line_chart function
def create_line_chart(data, title, xlabel, ylabel, color):
    """
    Create a line chart based in the data argument.

    Input:
    data = DataFrame to use for plotting the data
    title = Chart Title
    xlabel = Label for X Axis
    ylabel = Label for Y Axis
    color = Colour of the bar chart

    """
    data.plot(kind="line", xlabel=xlabel, ylabel=ylabel, color=color, title=title)

    #We want to show the plot after plotting it, if we dont do this, it will overwrite the
    plt.show()
```

In [10]:

```
# Create two line charts, one to plot the monthly shelter costs for owned dwelleing and ou

# Line chart for owned dwellings

create_line_chart(data=monthly_shelter_cost_per_year["shelter_costs_owned"], title="Averag
                  xlabel="Year", ylabel="Average Monthly Shelter Costs", color="red")
# Line chart for rented dwellings
create_line_chart(data=monthly_shelter_cost_per_year["shelter_costs_rented"], title="Avera
                  xlabel="Year", ylabel="Average Monthly Shelter Costs", color="yellow")
```





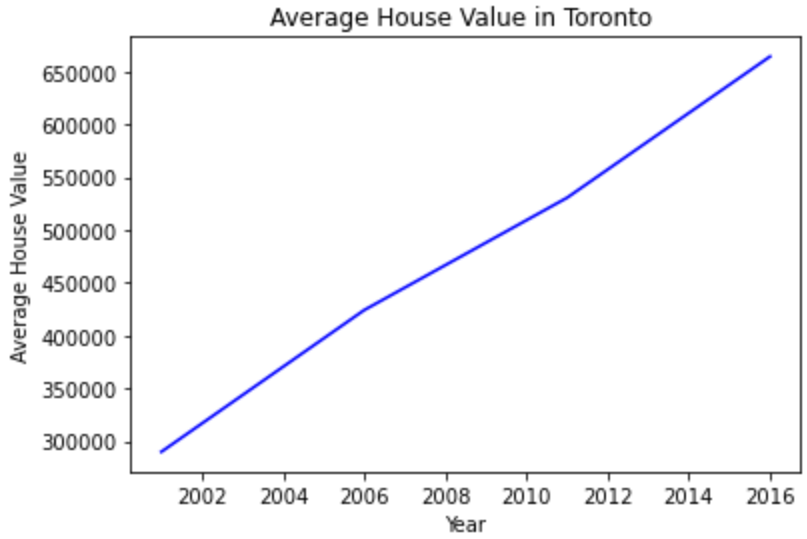
Average House Value per Year

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 [11]:
```
# Calculate the average house value per year
average_house_value = to_data["average_house_value"].groupby("year").mean()
average_house_value
```

Out[11]:
```
year
2001    289882.885714
2006    424059.664286
2011    530424.721429
2016    664068.328571
Name: average_house_value, dtype: float64
```

In [12]:
```
# Plot the average house value per year as a line chart
create_line_chart(data=average_house_value, title="Average House Value in Toronto ",
                  xlabel="Year", ylabel="Average House Value", color="blue")
```



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.

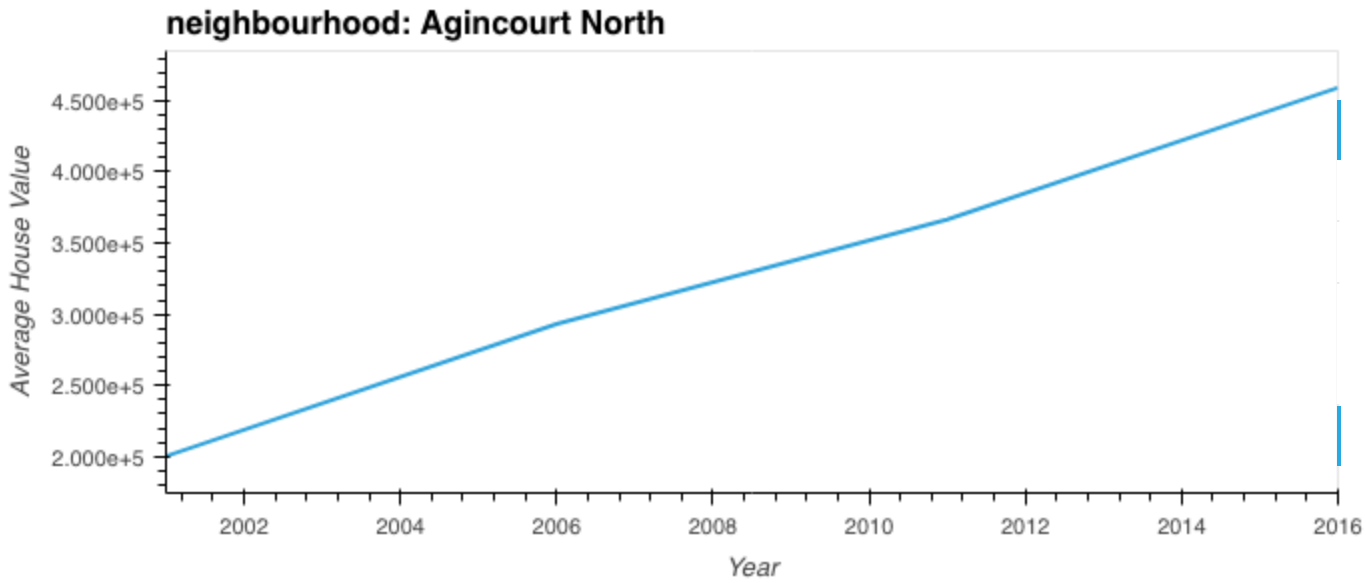
In [13]:
```
# Create a new DataFrame with the mean house values by neighbourhood per year
avg_house_value_by_neighbourhood = to_data[["neighbourhood","average_house_value"]].reset_
avg_house_value_by_neighbourhood.head(10)
```

Out[13]:

	year	neighbourhood	average_house_value
0	2001	Agincourt North	200388
1	2001	Agincourt South-Malvern West	203047
2	2001	Alderwood	259998
3	2001	Annex	453850
4	2001	Banbury-Don Mills	371864
5	2001	Bathurst Manor	304749
6	2001	Bay Street Corridor	257404
7	2001	Bayview Village	327644
8	2001	Bayview Woods-Steeles	343535
9	2001	Bedford Park-Nortown	565304

In [14]:
```
# Use hvplot to create an interactive line chart of the average house value per neighbourh
# The plot should have a dropdown selector for the neighbourhood
avg_house_value_by_neighbourhood.hvplot.line(x="year", y="average_house_value", xlabel="Ye
```

Out[14]:



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.

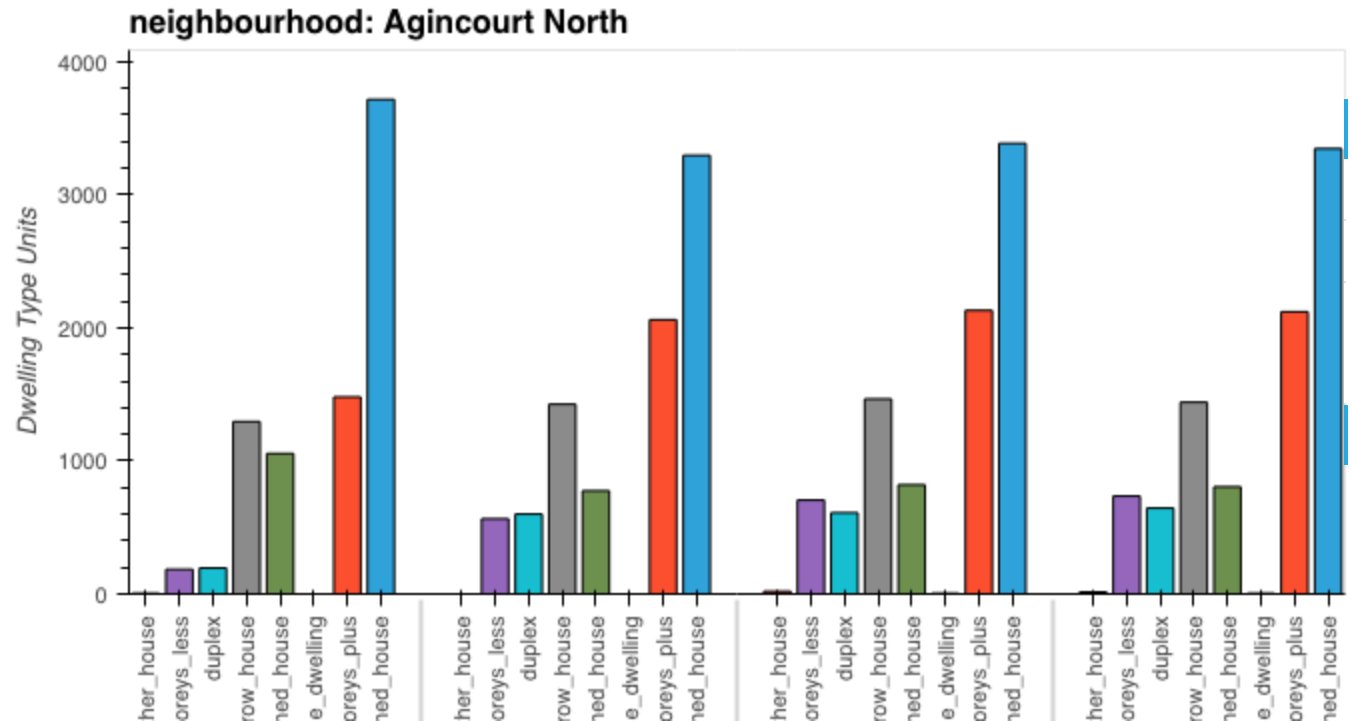
In [15]:
```
# Fetch the data of all dwelling types per year
dwelling_types_per_year = to_data.drop(columns=["average_house_value", "shelter_costs_owne
dwelling_types_per_year.head(10)
```

Out[15]:

year	neighbourhood	single_detached_house	apartment_five_storeys_plus	movable_dwelling	semi_detached
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	
2001	Bathurst Manor	2405	1550	0	
2001	Bay Street Corridor	10	7575	0	
2001	Bayview Village	2170	630	0	
2001	Bayview Woods-Steeles	1650	1715	0	
2001	Bedford Park-Nortown	4985	2080	0	

In [13]:
```
# Use hvplot to create an interactive bar chart of the number of dwelling types per neighb
# The plot should have a dropdown selector for the neighbourhood

dwelling_types_per_year.hvplot.bar(groupby="neighbourhood",rot=90, ylabel="Dwelling Type U
```

Out[13]:



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.

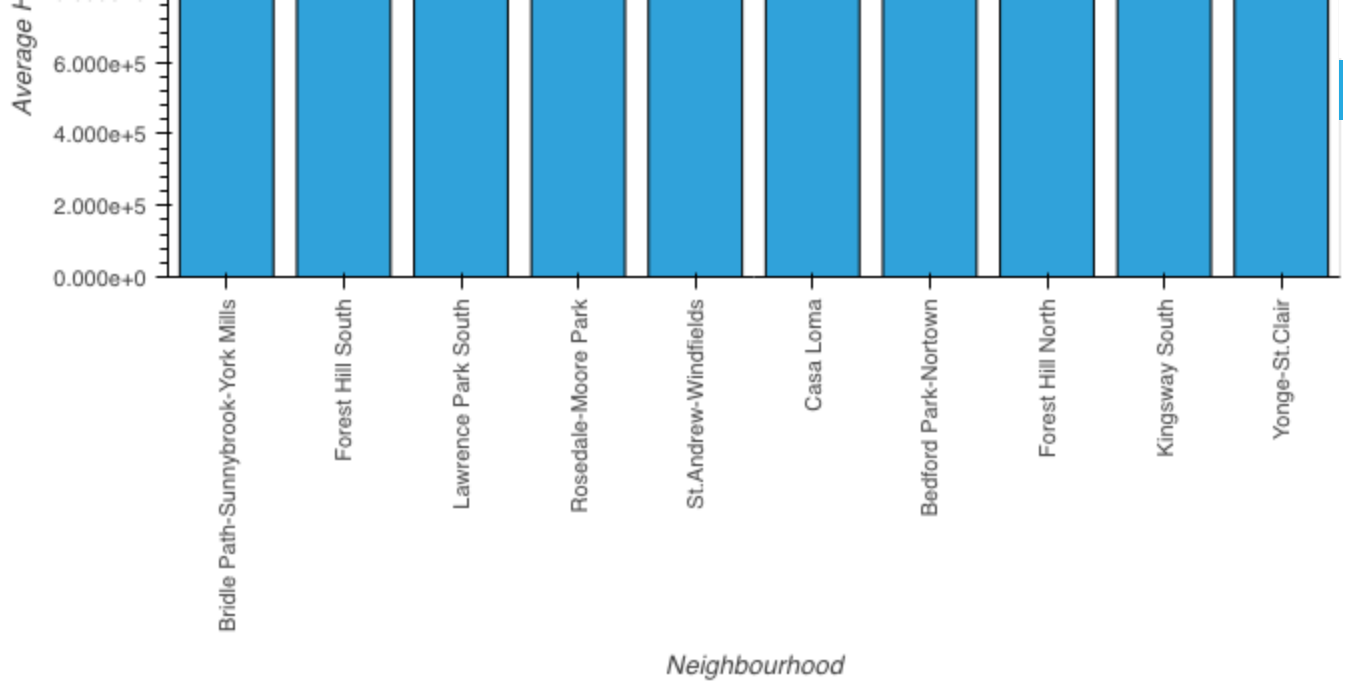
In [16]:
```
# Getting the data from the top 10 expensive neighbourhoods
ten_expensive_neighbourhoods = to_data.groupby("neighbourhood").mean().sort_values(by="ave
ten_expensive_neighbourhoods.head(10)
```

Out[16]:

	neighbourhood	single_detached_house	apartment_five_storeys_plus	movable_dwelling	semi_detached_h
0	Bridle Path-Sunnybrook-York Mills	2260.00	331.25	0.00	3
1	Forest Hill South	1742.50	2031.25	1.25	6
2	Lawrence Park South	3472.50	773.75	0.00	12
3	Rosedale-Moore Park	2498.75	4641.25	0.00	48
4	St.Andrew-Windfields	3225.00	1670.00	0.00	18
5	Casa Loma	916.25	2310.00	0.00	28
6	Bedford Park-Nortown	4865.00	1981.25	0.00	4
7	Forest Hill North	1488.75	3392.50	0.00	1
8	Kingsway South	2326.25	576.25	0.00	6
9	Yonge-St.Clair	565.00	3948.75	0.00	42

In [17]:
```
# Plotting the data from the top 10 expensive neighbourhoods
ten_expensive_neighbourhoods.head(10).hvplot.bar(rot=90, ylabel="Average House Value", xla
#ten_expensive_neighbourhoods[["average_house_value","shelter_costs_rented"]]
```

Out[17]:



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 [18]:

```
# Load neighbourhoods coordinates data
file_path = Path("Data/toronto_neighbourhoods_coordinates.csv")
df_neighbourhood_locations = pd.read_csv(file_path)
df_neighbourhood_locations
```

Out[18]:

	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
...	...	...	...
135	Wychwood	43.676919	-79.425515
136	Yonge-Eglinton	43.704689	-79.403590
137	Yonge-St.Clair	43.687859	-79.397871
138	York University Heights	43.765736	-79.488883
139	Yorkdale-Glen Park	43.714672	-79.457108

140 rows × 3 columns

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 [19]:
```
# Calculate the mean values for each neighborhood
mean_data_neighbourhoods = to_data.groupby("neighbourhood").mean().reset_index()
mean_data_neighbourhoods.head()
```

Out[19]:

	neighbourhood	single_detached_house	apartment_five_storeys_plus	movable_dwelling	semi_detached_ho
0	Agincourt North	3435.00	1947.50	2.50	86
1	Agincourt South-Malvern West	2897.50	2180.00	1.25	37
2	Alderwood	2903.75	302.50	1.25	50
3	Annex	751.25	7235.00	1.25	137
4	Banbury-Don Mills	3572.50	5388.75	1.25	27

In [20]:
```
#Set the index to neighbourhood so we join without any errors or inaccuracies
mean_data_neighbourhoods.set_index(keys="neighbourhood",inplace=True)
df_neighbourhood_locations.set_index(keys="neighbourhood",inplace=True)
```

In [21]:
```
# Join the average values with the neighbourhood locations
neighbourhood_with_location = pd.concat([mean_data_neighbourhoods, df_neighbourhood_locati
neighbourhood_with_location
```

Out[21]:

neighbourhood	single_detached_house	apartment_five_storeys_plus	movable_dwelling	semi_detached_house
Agincourt North	3435.00	1947.50	2.50	863.75
Agincourt South-Malvern West	2897.50	2180.00	1.25	375.00
Alderwood	2903.75	302.50	1.25	503.75
Annex	751.25	7235.00	1.25	1375.00
Banbury-Don Mills	3572.50	5388.75	1.25	273.75
...	...	...	...	...
Wychwood	1056.25	1236.25	0.00	992.50
Yonge-Eglinton	1468.75	1638.75	0.00	470.00
Yonge-St.Clair	565.00	3948.75	0.00	425.00
York University Heights	1355.00	5165.00	1.25	1316.25

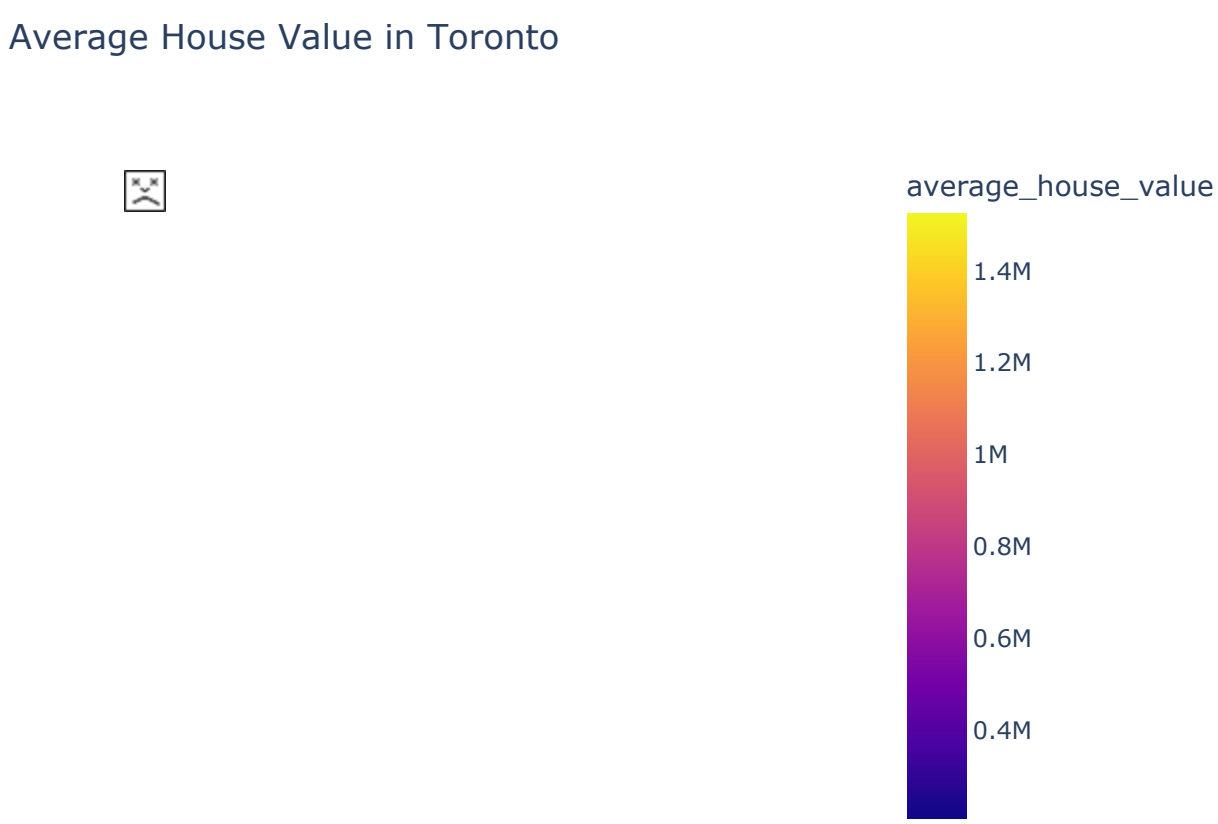
neighbourhood	single_detached_house	apartment_five_storeys_plus	movable_dwelling	semi_detached_hous
Yorkdale-Glen Park	2286.25	1347.50	0.00	73.7

140 rows × 13 columns

Mapbox Visualization

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

In [22]:
```
# Create a scatter mapbox to analyze neighbourhood info
avg_house_value_map= px.scatter_mapbox(
    neighbourhood_with_location,
    lat="lat",
    lon="lon",
color="average_house_value", title="Average House Value in Toronto"
)
avg_house_value_map.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 [23]:
```
mean_data_neighbourhoods.reset_index(inplace=True)
```

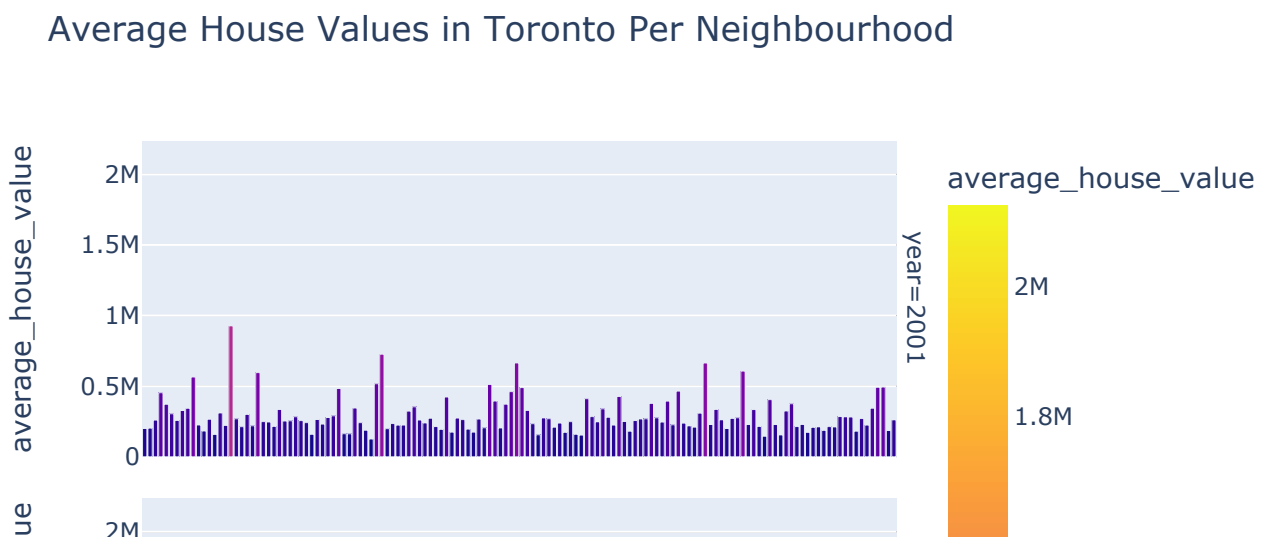
In [24]:
```
to_data
```

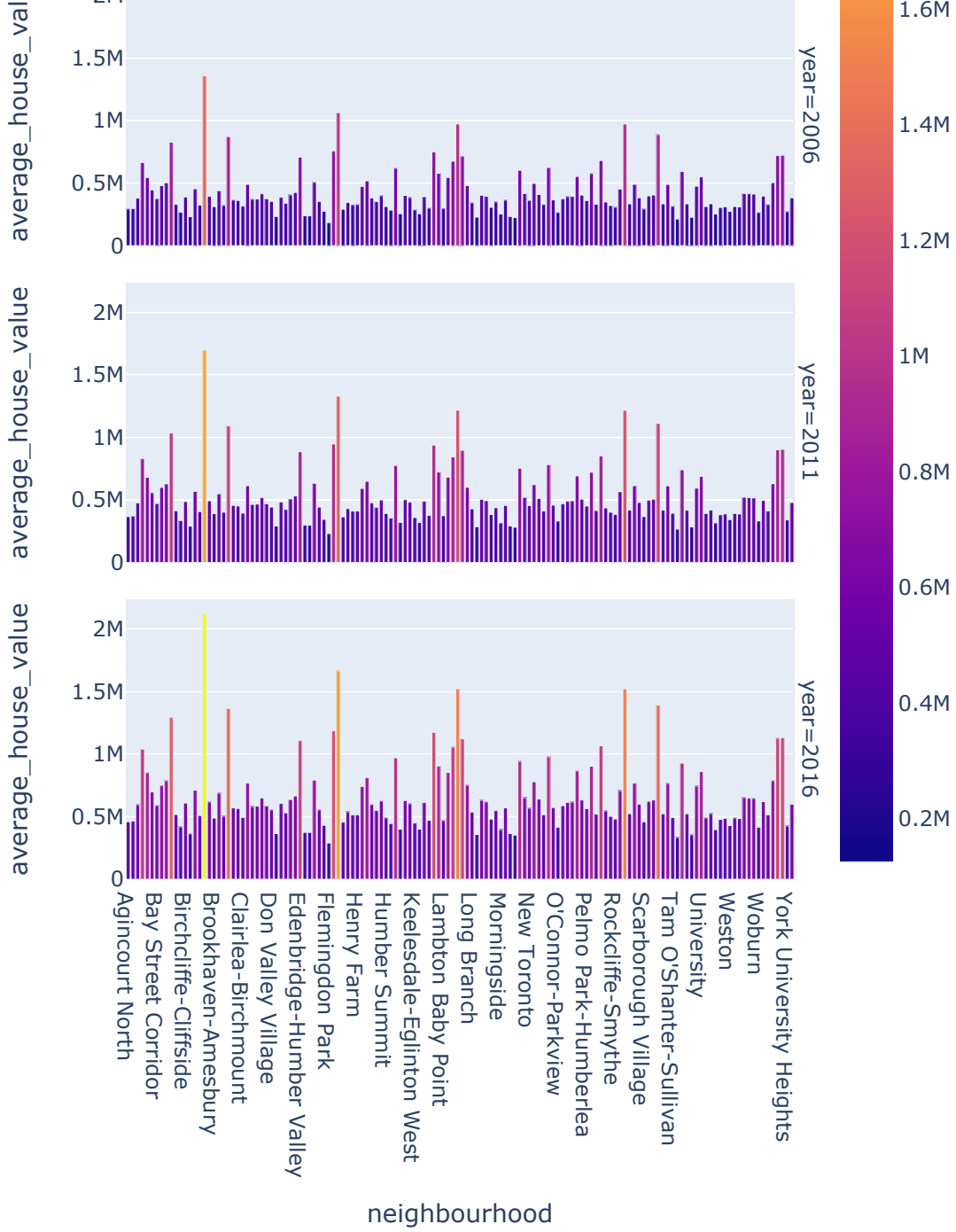
Out[24]:

year	neighbourhood	single_detached_house	apartment_five_storeys_plus	movable_dwelling	semi_detached
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	
...	...	...	...	...	
2016	Wychwood	920	1295	0	
2016	Yonge-Eglinton	1400	1995	0	
2016	Yonge-St.Clair	520	4315	0	
2016	York University Heights	1235	5505	0	
2016	Yorkdale-Glen Park	2165	1185	0	

560 rows × 12 columns

In [25]:
```
px.bar(to_data, x="neighbourhood", y="average_house_value",color="average_house_value", fa
```





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

In [26]:

```
# Fetch the data from all expensive neighbourhoods per year.
ten_most_expensive_neighbourhoods_per_year = to_data.sort_values(by="average_house_value",
ten_most_expensive_neighbourhoods_per_year
```

Out[26]:

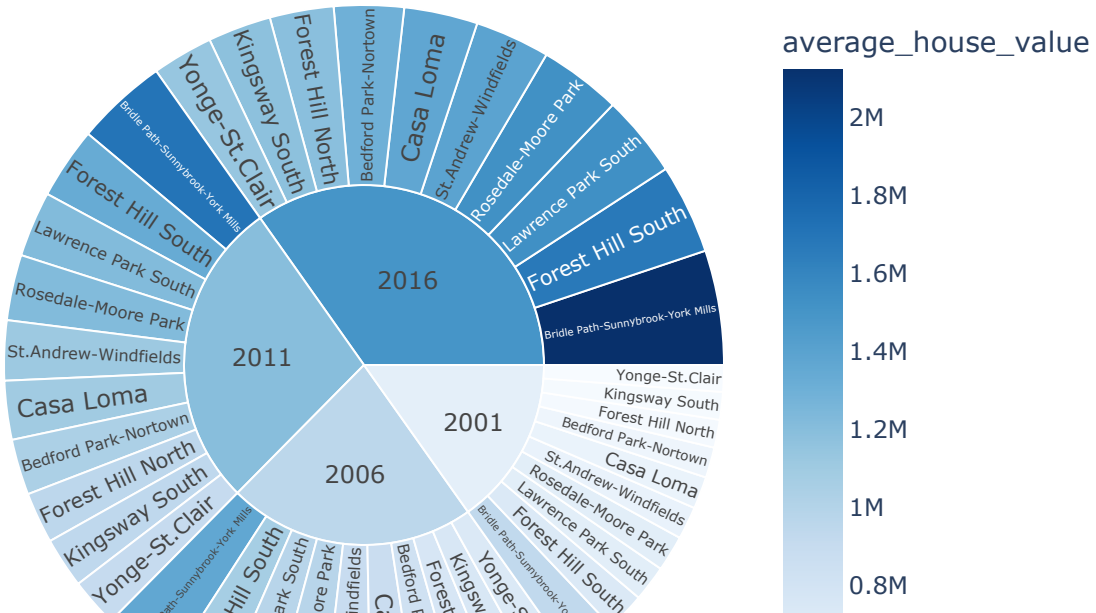
year	neighbourhood	single_detached_house	apartment_five_storeys_plus	movable_dwelling	semi_detache
2016	Bridle Path-Sunnybrook-York Mills	2275	590	0	
2011	Bridle Path-Sunnybrook-York Mills	2285	480	0	

year	neighbourhood	single_detached_house	apartment_five_storeys_plus	movable_dwelling	semi_detache
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	
2011	Lawrence Park South	3465	855	0	
2011	Rosedale-Moore Park	2485	4905	0	
2016	Forest Hill North	1470	3430	0	
2016	Kingsway South	2310	790	0	
2016	Yonge-St.Clair	520	4315	0	
2011	St.Andrew-Windfields	3285	1740	0	
2011	Casa Loma	880	2630	0	
2006	Forest Hill South	1740	1835	0	
2011	Bedford Park-Nortown	4870	1960	0	
2006	Lawrence Park South	3415	745	0	
2006	Rosedale-Moore Park	2450	4790	0	
2011	Forest Hill North	1470	3350	0	
2011	Kingsway South	2350	695	0	
2001	Bridle Path-Sunnybrook-York Mills	2275	110	0	
2011	Yonge-St.Clair	530	4070	0	
2006	St.Andrew-Windfields	3095	1740	0	
2006	Casa Loma	875	2230	0	

year	neighbourhood	single_detached_house	apartment_five_storeys_plus	movable_dwelling	semi_detache
2006	Bedford Park-Nortown	4785	1890	0	
2006	Forest Hill North	1450	3410	0	
2006	Kingsway South	2275	560	0	
2001	Forest Hill South	1815	2440	5	
2006	Yonge-St.Clair	540	3785	0	
2001	Lawrence Park South	3590	570	0	
2001	Rosedale-Moore Park	2610	3880	0	
2001	St.Andrew-Windfields	3275	1455	0	
2001	Casa Loma	1035	1700	0	
2001	Bedford Park-Nortown	4985	2080	0	
2001	Forest Hill North	1565	3380	0	
2001	Kingsway South	2370	260	0	
2001	Yonge-St.Clair	670	3625	0	

In [27]:

```
# Create the sunburst chart
px.sunburst(
    ten_most_expensive_neighbourhoods_per_year,
    path=[ten_most_expensive_neighbourhoods_per_year.index, 'neighbourhood'], values='aver
    height=500

)
```




Forest
Lawrence
Rosedale-Mo
St.Andrew-W
asa Loma
park-Nortown
Hill North
ay South
St.Clair
0.6M