Assignment_09: Housing Price Analysis

In [1]:
```
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sns
from sklearn.preprocessing import StandardScaler
from sklearn.pipeline import Pipeline
from sklearn.linear_model import LinearRegression
import sklearn.metrics as metrics
from sklearn.model_selection import train_test_split
from sklearn.metrics import r2_score
```

In [8]:
```
df = pd.read_csv('Housing Data.csv')
```

In [9]:
```
df.head()
```

Out[9]:

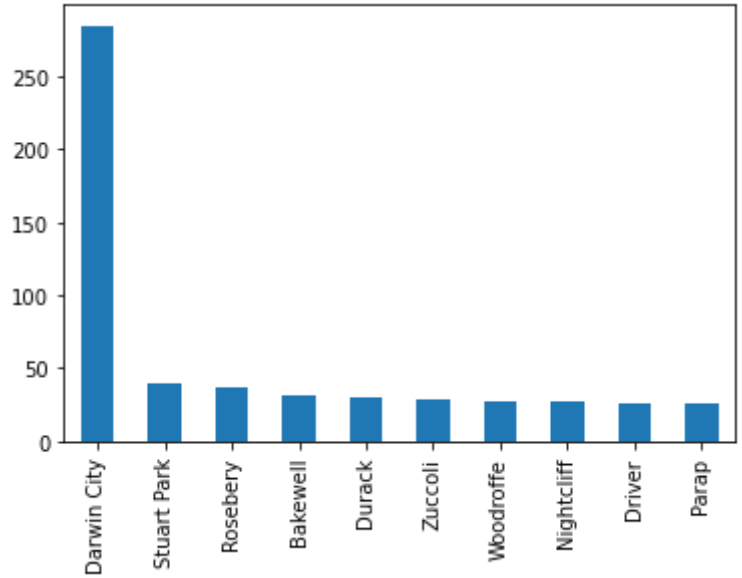
	index	TID	breadcrumb	category_name	property_type	building_size	land_size	preferred_size	open_date	listing_agency	...	stat
0	0	1350988	Buy>NT>DARWIN CITY	Real Estate & Property for sale in DARWIN CITY...	House	NaN	NaN	NaN	Added 2 hours ago	Professionals - DARWIN CITY	...	N
1	1	1350989	Buy>NT>DARWIN CITY	Real Estate & Property for sale in DARWIN CITY...	Apartment	171m²	NaN	171m²	Added 7 hours ago	Nick Mousellis Real Estate - Eview Group Member	...	N
2	2	1350990	Buy>NT>DARWIN CITY	Real Estate & Property for sale in DARWIN CITY...	Unit	NaN	NaN	NaN	Added 22 hours ago	Habitat Real Estate - THE GARDENS	...	N
3	3	1350991	Buy>NT>DARWIN CITY	Real Estate & Property for sale in DARWIN CITY...	House	NaN	NaN	NaN	Added yesterday	Ray White - NIGHTCLIFF	...	N
4	4	1350992	Buy>NT>DARWIN CITY	Real Estate & Property for sale in DARWIN CITY...	Unit	201m²	NaN	201m²	Added yesterday	Carol Need Real Estate - Fannie Bay	...	N

5 rows × 27 columns

Q1: To find Relationships Between Housing prices and Location.

In [10]:

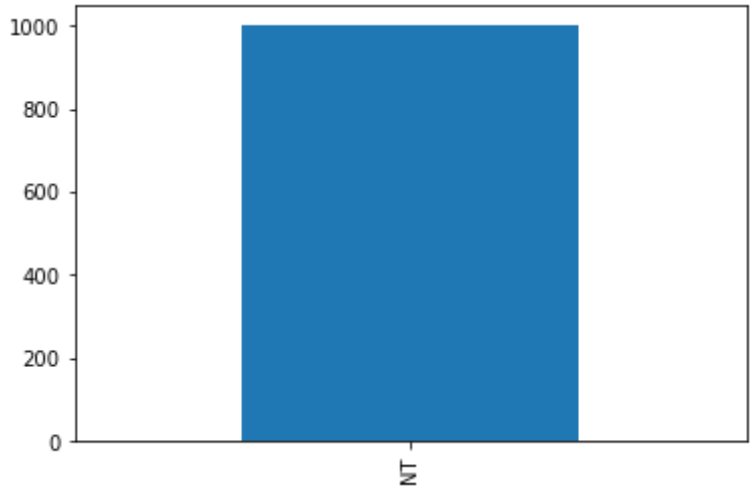
```
city =  df['city'].value_counts()
city.head(10).plot.bar()
plt.show()
```



Observation: In terms of City Darwin City has More Demand

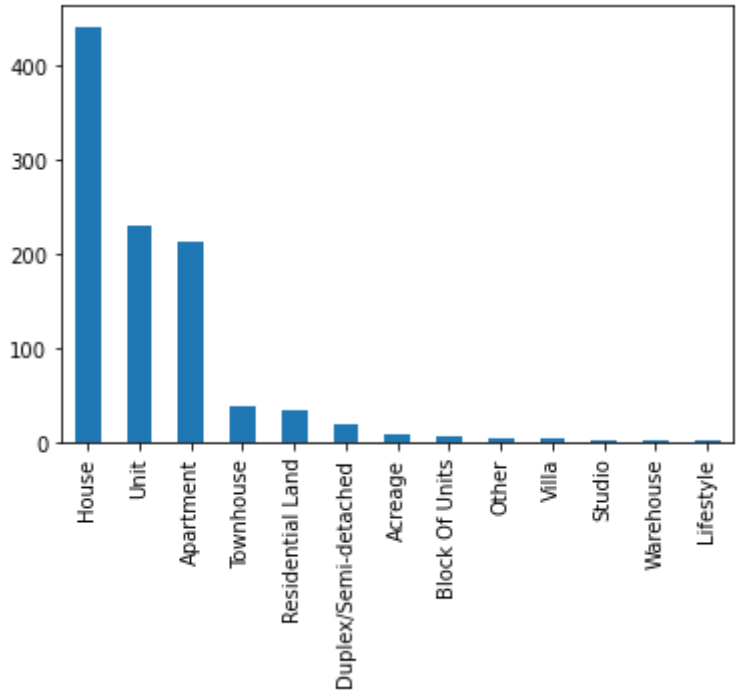
In [11]:

```
state =  df['state'].value_counts()
state.head().plot.bar()

plt.show()
```



In [12]:

```
protype =  df['property_type'].value_counts()
protype.head(20).plot.bar()

plt.show()
```



Observation: People were interested in House Purchase and then Unit and Apartment

Q2: To find relationships between housing prices and features such as size, number of bedrooms, and number of bathrooms

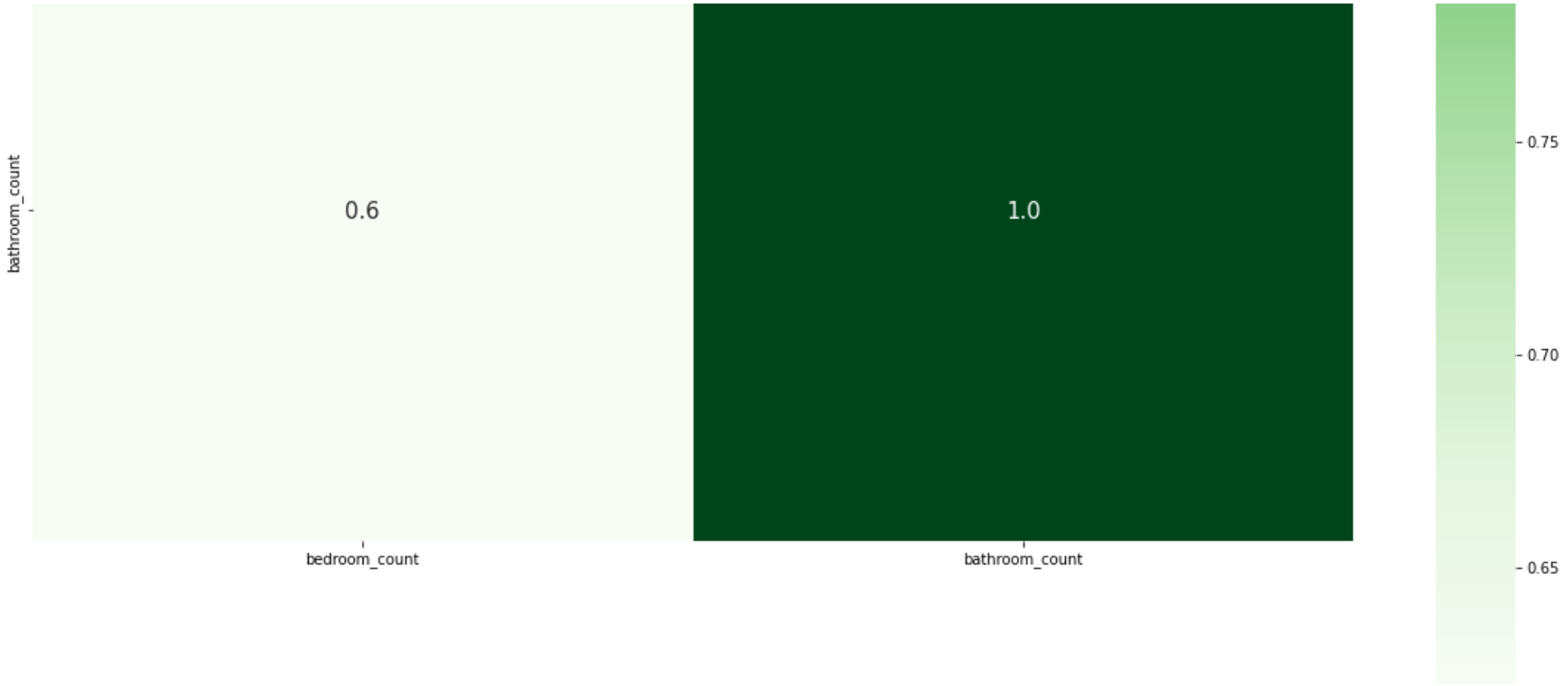
In [13]:

```
df1 = df[['price', 'bedroom_count', 'bathroom_count']].copy()
df1 = df1.dropna()
```

In [15]:

```
corr = df1.corr()
corr.shape
# Plotting the heatmap of correlation between features
plt.figure(figsize=(20,20))

sns.heatmap(corr, cbar=True, square= True, fmt='.1f', annot=True, annot_kws={'size':15}, cmap='Greens')
```

Out[15]: <AxesSubplot:>



In [17]:

```
df['building_size'] = df['building_size'].str.replace(r'\D', '')
df['building_size'] = pd.to_numeric(df['building_size'])
```

```
C:\Users\I5262\AppData\Local\Temp\ipykernel_13064\2222709904.py:1: FutureWarning: The default value of regex will c
hange from True to False in a future version.
  df['building_size'] = df['building_size'].str.replace(r'\D', '')
```

In [18]:
```
boxplot = df.boxplot(column=['building_size'])
```