

Analysis of Development Projects in Louisville, Kentucky

The dataset that I have selected for the purpose of this project is about the applications for development projects submitted to Planning and Design Services in the city of Louisville, Kentucky between January 2010 and June 2018.

I have used the tools R, Python and SQL to analyze and provide inferences and draw conclusions from the dataset.

Research questions that I have analyzed this dataset and answered using the above-mentioned programming tools:

1. How many number of projects are there per Project type?

Approach 1: Python

```
Out[180]:
```

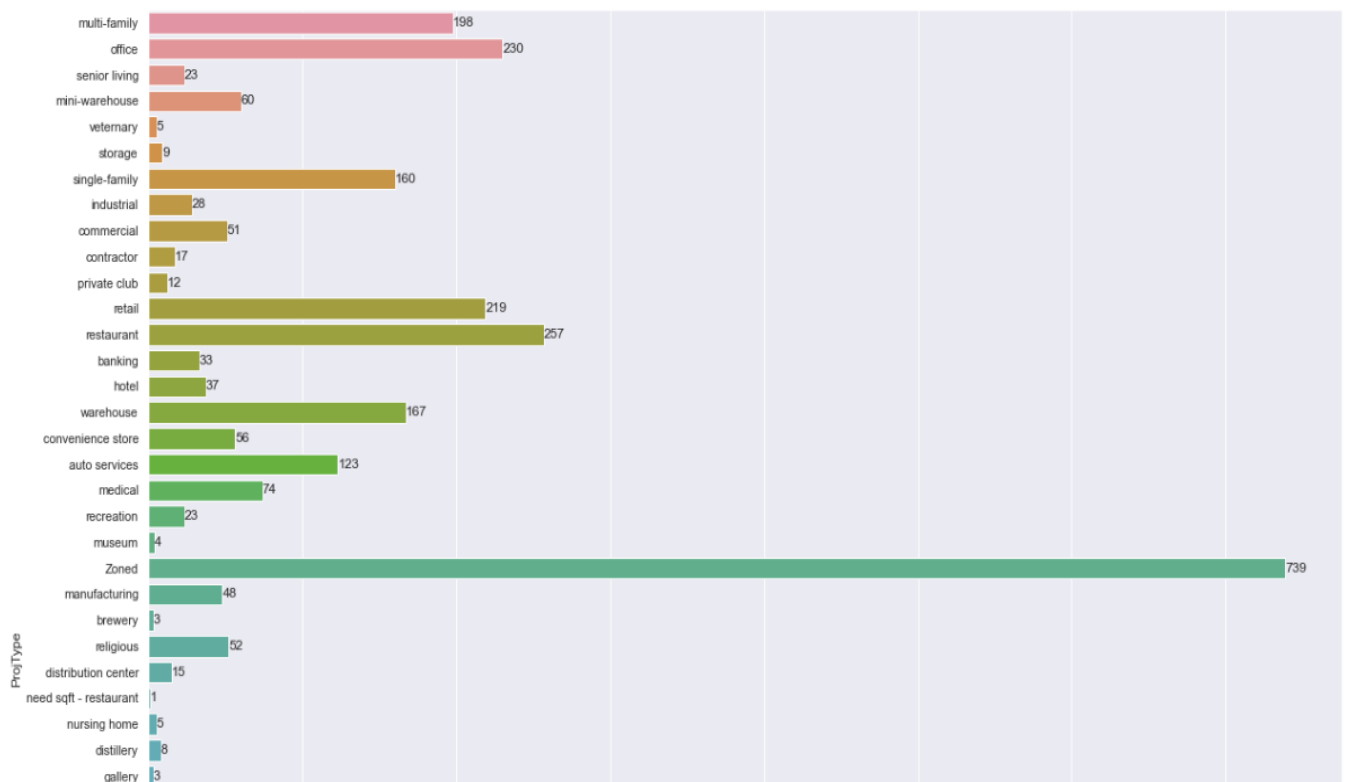
| | |
|---------------------|-----|
| Zoned | 739 |
| restaurant | 257 |
| office | 230 |
| retail | 219 |
| multi-family | 198 |
| warehouse | 167 |
| single-family | 160 |
| auto services | 123 |
| medical | 74 |
| mini-warehouse | 60 |
| convenience store | 56 |
| religious | 52 |
| commercial | 51 |
| manufacturing | 48 |
| hotel | 37 |
| banking | 33 |
| industrial | 28 |
| senior living | 23 |
| recreation | 23 |
| grocery | 19 |
| contractor | 17 |
| distribution center | 15 |
| daycare | 13 |
| assisted living | 13 |
| private club | 12 |
| fuel kiosk | 10 |
| storage | 9 |
| distillery | 8 |
| funeral home | 7 |
| veterinary | 5 |
| nursing home | 5 |
| museum | 4 |
| conference center | 4 |

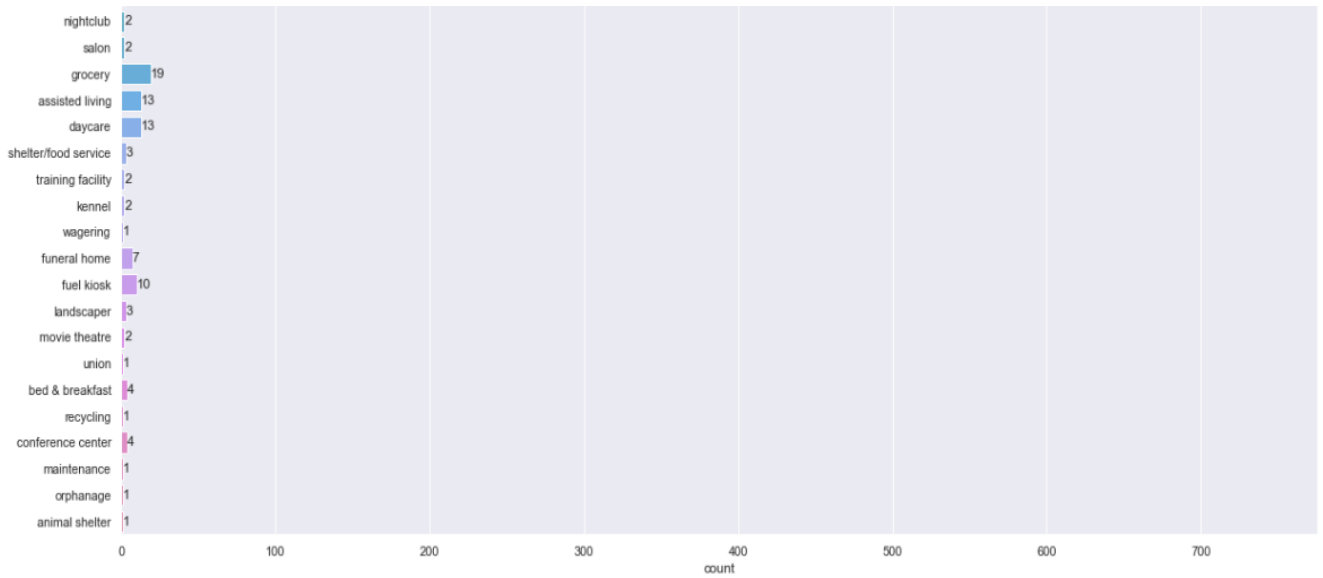
| | |
|------------------------|---|
| bed & breakfast | 4 |
| gallery | 3 |
| shelter/food service | 3 |
| brewery | 3 |
| landscaper | 3 |
| movie theatre | 2 |
| training facility | 2 |
| kennel | 2 |
| salon | 2 |
| nightclub | 2 |
| wagering | 1 |
| union | 1 |
| recycling | 1 |
| need sqft - restaurant | 1 |
| maintenance | 1 |
| orphanage | 1 |
| animal shelter | 1 |

Name: ProjType, dtype: int64

From the above counts, it can be seen that the Zoned project type has the highest number of projects, at 739, followed by Restaurants at 257.

Visualization:





Approach 2: R

```
> table(development_data$ProjType)
```

| | | | | |
|----------------|----------------------|---------------------|------------------------|-------------------|
| animal shelter | assisted living | auto services | banking | bed & breakfast |
| 1 | 13 | 123 | 33 | 4 |
| brewery | commercial | conference center | contractor | convenience store |
| 3 | 51 | 4 | 17 | 56 |
| daycare | distillery | distribution center | fuel kiosk | funeral home |
| 13 | 8 | 15 | 10 | 7 |
| gallery | grocery | hotel | industrial | kennel |
| 3 | 19 | 37 | 28 | 2 |
| landscaper | maintenance | manufacturing | medical | mini-warehouse |
| 3 | 1 | 48 | 74 | 60 |
| movie theatre | multi-family | museum | need sqft - restaurant | nightclub |
| 2 | 198 | 4 | 1 | 2 |
| nursing home | office | orphanage | private club | recreation |
| 5 | 230 | 1 | 12 | 23 |
| recycling | religious | restaurant | retail | salon |
| 1 | 52 | 257 | 219 | 2 |
| senior living | shelter/food service | single-family | storage | training facility |
| 23 | 3 | 160 | 9 | 2 |
| union | veterinary | wagering | warehouse | Zoned |
| 1 | 5 | 1 | 167 | 739 |

Approach 3: SQL

| PROJ_TYPE | NUMBER_OF_PROJECTS |
|---------------------------|--------------------|
| 1 Zoned | 739 |
| 2 restaurant | 257 |
| 3 office | 230 |
| 4 retail | 219 |
| 5 multi-family | 198 |
| 6 warehouse | 167 |
| 7 single-family | 160 |
| 8 auto services | 123 |
| 9 medical | 74 |
| 10 mini-warehouse | 60 |
| 11 convenience store | 56 |
| 12 religious | 52 |
| 13 commercial | 51 |
| 14 manufacturing | 48 |
| 15 hotel | 37 |
| 16 banking | 33 |
| 17 industrial | 28 |
| 18 senior living | 23 |
| 19 recreation | 23 |
| 20 grocery | 19 |
| 21 contractor | 17 |
| 22 distribution center | 15 |
| 23 assisted living | 13 |
| 24 daycare | 13 |
| 25 private club | 12 |
| 26 fuel kiosk | 10 |
| 27 storage | 9 |
| 28 distillery | 8 |
| 29 funeral home | 7 |
| 30 veterinary | 5 |
| 31 nursing home | 5 |
| 32 museum | 4 |
| 33 bed & breakfast | 4 |
| 34 conference center | 4 |
| 35 gallery | 3 |
| 36 brewery | 3 |
| 37 shelter/food service | 3 |
| 38 landscaper | 3 |
| 39 nightclub | 2 |
| 40 salon | 2 |
| 41 training facility | 2 |
| 42 kennel | 2 |
| 43 movie theatre | 2 |
| 44 need sqft - restaurant | 1 |
| 45 wagering | 1 |
| 46 union | 1 |
| 47 recycling | 1 |
| 48 maintenance | 1 |
| 49 orphanage | 1 |
| 50 animal shelter | 1 |

2. Identify the number of projects per Application Type and Category.

Approach 1 : Python

2. Identify the number of projects per Application Type and Category.

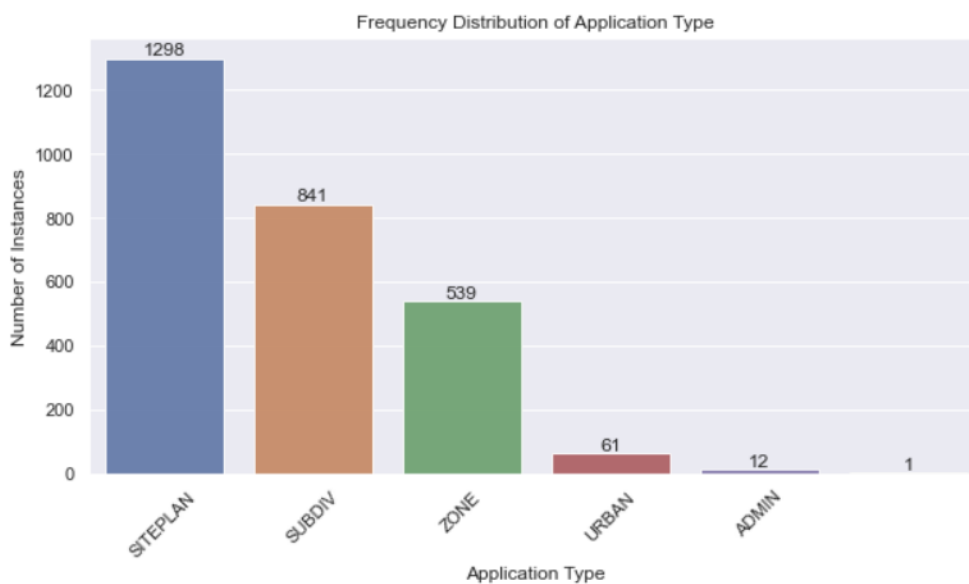
```
pd.set_option("display.max_rows", None)|  
development_subset.AppType.value_counts()
```

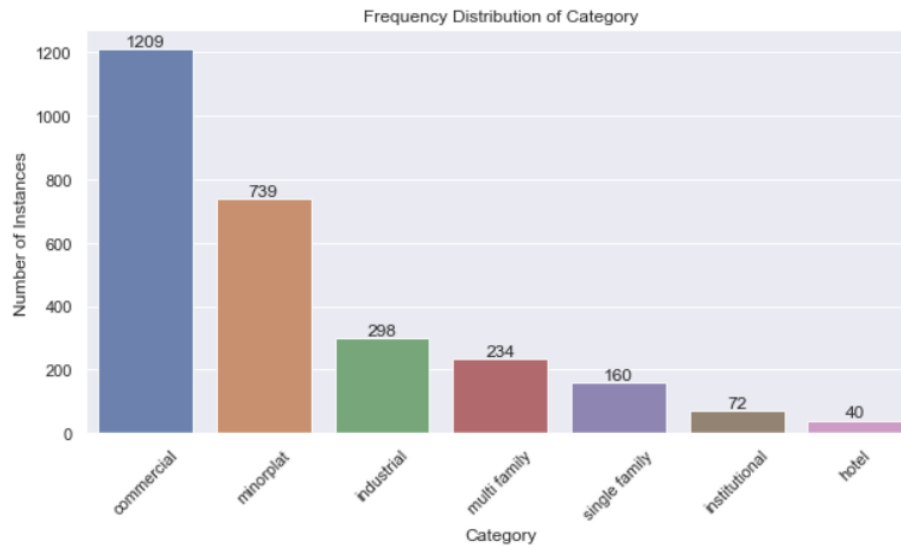
```
SITEPLAN      1298  
SUBDIV        841  
ZONE          539  
URBAN         61  
ADMIN         12  
              1  
Name: AppType, dtype: int64
```

```
pd.set_option("display.max_rows", None)  
development_subset.Category.value_counts()
```

```
commercial      1209  
minorplat       739  
industrial      298  
multi family    234  
single family   160  
institutional   72  
hotel           40  
Name: Category, dtype: int64
```

Visualization:





Approach 2: SQL

| | APP_TYPE | PROJ_TYPE | NUMBER_OF_PROJECTS |
|---|----------|---------------|--------------------|
| 1 | SITEPLAN | commercial | 1298 |
| 2 | SUBDIV | single-family | 841 |
| 3 | ZONE | multi-family | 539 |
| 4 | URBAN | commercial | 61 |
| 5 | ADMIN | restaurant | 12 |
| 6 | | retail | 1 |

| | CATEGORY_ | PROJ_TYPE | NUMBER_OF_PROJECTS |
|---|---------------|---------------|--------------------|
| 1 | commercial | office | 1209 |
| 2 | minorplat | Zoned | 739 |
| 3 | industrial | storage | 298 |
| 4 | multi family | multi-family | 234 |
| 5 | single family | single-family | 160 |
| 6 | institutional | recreation | 72 |
| 7 | hotel | hotel | 40 |

3. Which category and Application Type have the highest number of projects?

Approach 1: Python

```
# 3. Which category and Application Type have the highest number of projects?
```

```
pd.set_option("display.max_rows", None)
development_subset.AppType.value_counts().head(1)
```

```
SITEPLAN    1298
Name: AppType, dtype: int64
```

```
pd.set_option("display.max_rows", None)
development_subset.Category.value_counts().head(1)
```

```
commercial    1209
Name: Category, dtype: int64
```

Approach 2: SQL

| | APP_TYPE | PROJ_TYPE | NUMBER_OF_PROJECTS |
|---|----------|------------|--------------------|
| 1 | SITEPLAN | commercial | 1298 |

| | CATEGORY_ | PROJ_TYPE | NUMBER_OF_PROJECTS |
|---|------------|-----------|--------------------|
| 1 | commercial | office | 1209 |

4. List and count of projects that are hospitals.

Approach 1: Python

| | SqFootage | ProjType | Rooms | AppType | ProjName | Category | CreYear | LastYear | BOZA | PC | LDT | DRC | Units |
|------|-----------|------------|-------|----------|--------------------------------|------------|---------|----------|----------|----|------------|----------|-------|
| 770 | 172263 | medical | 0 | ZONE | Norton Brownsboro Hospital | commercial | 2017 | 2018 | Approved | | | | 0 |
| 783 | 13842 | medical | 0 | ZONE | Norton Children's Hospital | commercial | 2017 | 2017 | Approved | | | | 0 |
| 1377 | 86266 | medical | 0 | SITEPLAN | Hospital expansion | commercial | 2015 | 2016 | Approved | | | Approved | 0 |
| 1415 | 2200 | medical | 0 | SITEPLAN | Norton Brownsboro Hospital | commercial | 2015 | 2015 | Approved | | | | 0 |
| 1555 | 3235 | office | 0 | ZONE | American Hospital Directory | commercial | 2014 | 2014 | Approved | | Sent to PC | | 0 |
| 2165 | 7000 | medical | 0 | ZONE | BAPTIST HOSPITAL EAST CANCER C | commercial | 2012 | 2012 | Approved | | | | 0 |
| 2225 | 120000 | medical | 0 | ZONE | NORTON WOMEN'S HOSPITAL KOSAIR | commercial | 2012 | 2012 | Approved | | | | 0 |
| 2404 | 0 | Zoned | 0 | SUBDIV | JEWISH HOSPITAL @ ST MARY'S HE | minorplat | 2011 | 2012 | | | | | 0 |
| 2647 | 2404 | veterinary | 0 | SITEPLAN | JEFFERSON ANIMAL HOSPITAL | commercial | 2010 | 2010 | Approved | | | Approved | 0 |

Approach 2: SQL

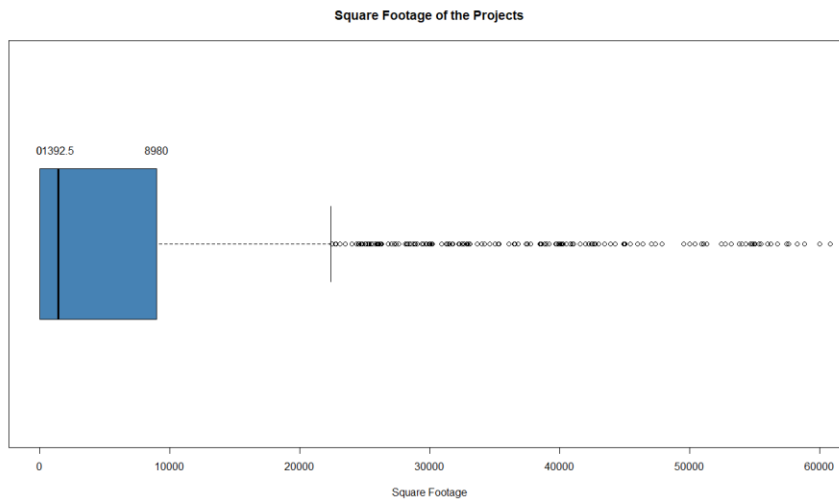
| PROJ_NAME |
|----------------------------------|
| 1 Norton Brownsboro Hospital |
| 2 Norton Children's Hospital |
| 3 Hospital expansion |
| 4 Norton Brownsboro Hospital |
| 5 American Hospital Directory |
| 6 BAPTIST HOSPITAL EAST CANCER C |
| 7 NORTON WOMEN'S HOSPITAL KOSAIR |
| 8 JEWISH HOSPITAL @ ST MARY'S HE |
| 9 JEFFERSON ANIMAL HOSPITAL |

| COUNT(PROJ_NAME) |
|------------------|
| 1 9 |

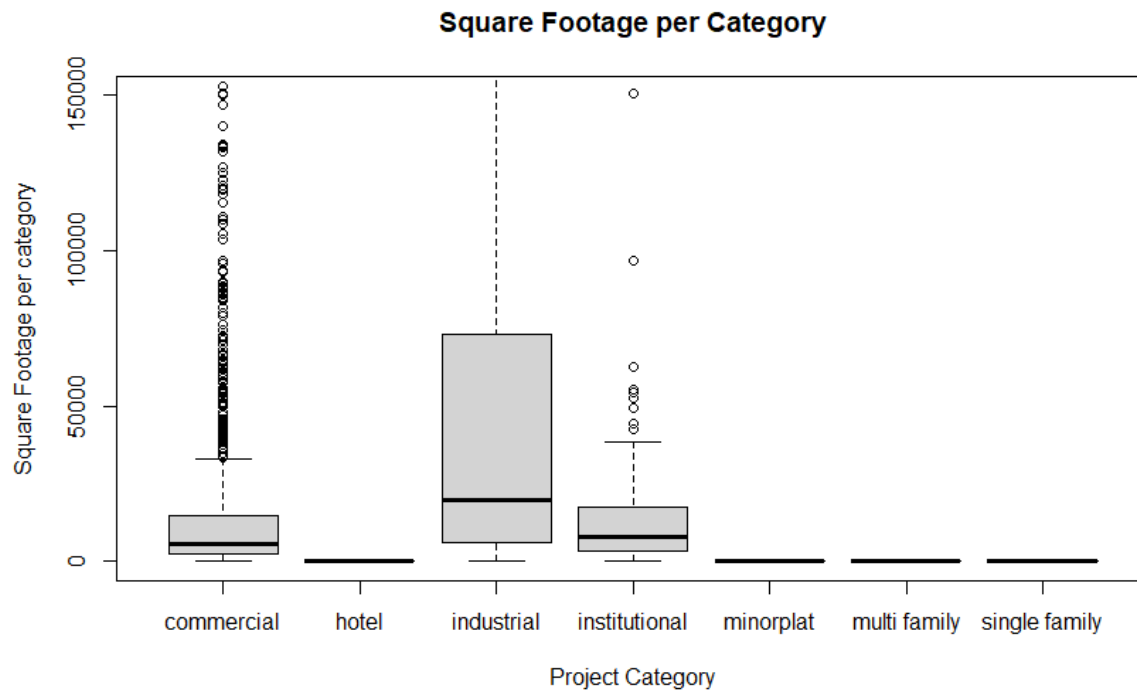
5. Identify the range for the Square Footage (area) of projects and Square Footage (area) per category. Also identify the summary statistics for the Square Footage (area) of projects.

Approach 1: R

```
> SqFootage_range = range(development_data$SqFootage, na.rm = TRUE)
> SqFootage_range
[1]      0 2555000
> |
>
> SqFootage_summary = summary(development_data$SqFootage, na.rm = TRUE)
> SqFootage_summary
  Min. 1st Qu.  Median    Mean 3rd Qu.    Max.
    0      0    1392   19768    8980 2555000
> |
```



Square Footage per Category:



From the above boxplots of Sq Footage per Category, it can be inferred that in general, Industrial projects have the highest area among all the categories of projects, followed by Commercial Projects.

6. Which project has highest square footage and highest number of rooms?

```
# 6. Which project has highest square footage and highest number of rooms?
```

```
Area_max = development_subset['SqFootage'].max()
development_subset.loc[development_subset['SqFootage'] == Area_max]
```

| | SqFootage | ProjType | Rooms | AppType | ProjName | Category | CreYear | LastYear | BOZA | PC | LDT | DRC | Units |
|------|-----------|-----------|-------|---------|----------------------|------------|---------|----------|------|----------|------------|-----|-------|
| 1837 | 2555000 | warehouse | 0 | ZONE | Southport Warehouses | industrial | 2013 | 2014 | | Approved | Sent to PC | | 0 |

```
Rooms_max = development_subset['Rooms'].max()
development_subset.loc[development_subset['Rooms'] == Rooms_max]
```

| | SqFootage | ProjType | Rooms | AppType | ProjName | Category | CreYear | LastYear | BOZA | PC | LDT | DRC | Units |
|------|-----------|----------|-------|----------|-----------------------|----------|---------|----------|----------|----|-----|-----|-------|
| 1456 | 0 | hotel | 600 | SITEPLAN | Omni Louisville Hotel | hotel | 2015 | 2015 | Approved | | | | 0 |

A warehouse project with the name of ‘Southport Warehouses’ has the highest area among all the projects, with an area of 2555000 sq. feet. This project belongs to the ‘Industrial’ Projects Category.

A Hotel with the name of ‘Omni Louisville Hotel’ has the highest number of rooms among all the projects, with 600 rooms. This project belongs to the ‘Hotel Category’.

Approach 2: SQL

| PROJ_NAME | SQ_FOOTAGE |
|------------------------|------------|
| 1 Southport Warehouses | 2555000 |

| PROJ_NAME | ROOMS |
|-------------------------|-------|
| 1 Omni Louisville Hotel | 600 |

7. How many projects are proposed to be built on a highway?

Approach 1: Python

```
# 7. How many projects are proposed to be built on a highway?
```

```
development_subset[development_subset['ProjName'].str.contains("highway | highway | highway ",case=False)]
```

| | SqFootage | ProjType | Rooms | AppType | ProjName | Category | CreYear | LastYear | BOZA | PC | LDT | DRC | Units |
|------|-----------|---------------|-------|----------|-------------------------------|------------|---------|----------|-------------|-------------|------------|-------------|-------|
| 35 | 1104 | office | 0 | ZONE | 4738 Dixie Highway | commercial | 2018 | 2018 | No decision | No decision | Sent to PC | No decision | 0 |
| 973 | 2300 | retail | 0 | SITEPLAN | 4724 Dixie Highway | commercial | 2016 | 2016 | | | | | 0 |
| 1032 | 3287 | grocery | 0 | SITEPLAN | Aldi #38 Preston Highway | commercial | 2016 | 2016 | | | | Approved | 0 |
| 1352 | 0 | Zoned | 0 | SUBDIV | 6650 Dixie Highway | minorplat | 2015 | 2015 | | | | | 0 |
| 1401 | 7373 | auto services | 0 | SITEPLAN | Discount Tire, Dixie Highway | commercial | 2015 | 2015 | | | | Approved | 0 |
| 2427 | 8000 | retail | 0 | SITEPLAN | FAMILY DOLLAR DIXIE HIGHWAY & | commercial | 2011 | 2011 | | | | Approved | 0 |

It can be seen that there are 6 projects that are proposed to be built on highways.

Approach 2: SQL

| COUNT(PROJ_NAME) | |
|------------------|---|
| 1 | 6 |

8. How many projects have been actioned upon for more than a year, since the initial application date?

Approach 1: R

```
> development_data$duration <- development_data$LastYear - development_data$CreYear
> names(development_data)
[1] "x" "SqFootage" "ProjType" "Rooms" "AppType" "ProjName" "Category" "CreYear" "LastYear"
[10] "BOZA" "PC" "LDT" "DRC" "Units" "duration"
> |
```

The proportion of projects that have been active for 0, 1, 2, 3 and 4 years:

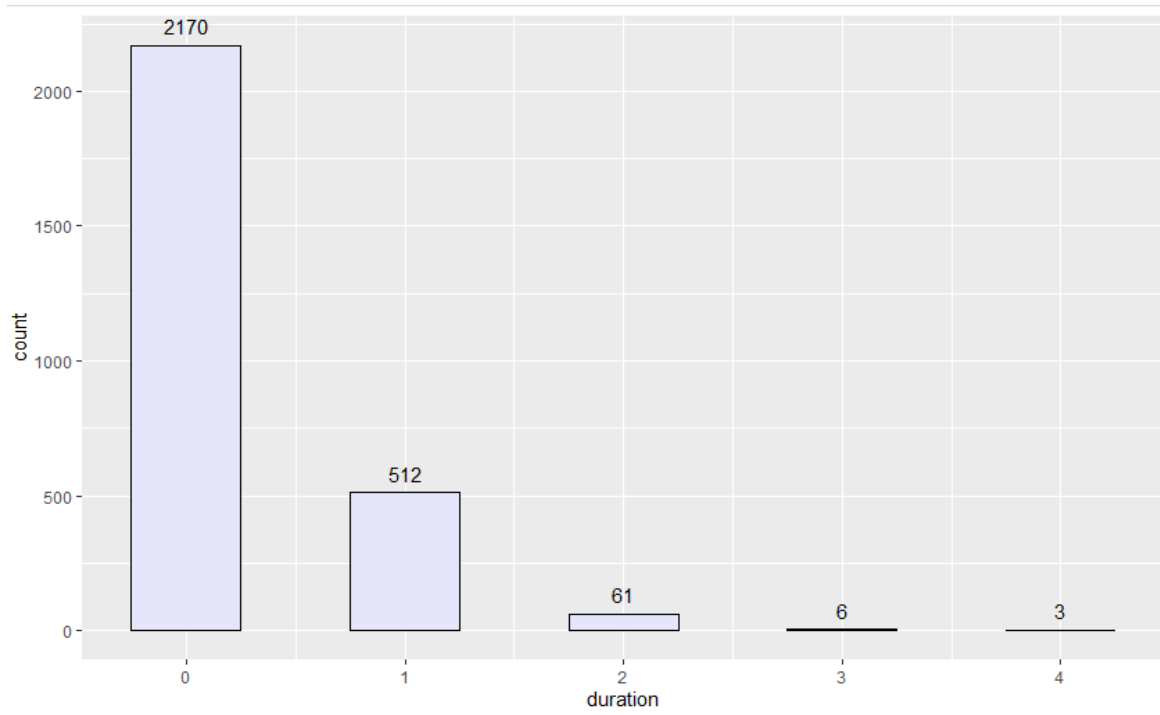
```
> prop.table(table(development_data$duration))
      0      1      2      3      4
0.788517442 0.186046512 0.022165698 0.002180233 0.001090116
> |
```

It shows that majority of the projects (78.8 %) have not been actioned upon for more than a year and 18.6 % of the projects have been active for 1 year.

The individual counts of the projects, grouped by the number of active years:

```
> sum(development_data$duration == 0)
[1] 2170
> sum(development_data$duration == 1)
[1] 512
> sum(development_data$duration == 2)
[1] 61
> sum(development_data$duration == 3)
[1] 6
> sum(development_data$duration == 4)
[1] 3
> |
```

Visual representation of the individual counts of the projects, grouped by the number of active years:



9. What proportion of projects have received the required approval of construction from the appropriate Boards/ Committees?

Approach 1: SQL

| | COUNT(*) |
|---|----------|
| 1 | 2752 |

BOZA:

| | BOZA | COUNT(BOZA) |
|---|-------------|-------------|
| 1 | No decision | 2417 |
| 2 | Approved | 335 |

Of the 2752 total projects, 335 applications have received the approval from BOZA (Board of Zoning Adjustment)

PC:

| PC | COUNT(PC) |
|---------------------|-----------|
| 1 No decision | 2355 |
| 2 Approved | 389 |
| 3 No Recommendation | 1 |
| 4 Continued | 3 |
| 5 Agreed with DRC | 3 |
| 6 Agreed with LD&T | 1 |

Of the 2752 total projects, 389 applications have received the approval from PC (Planning Commission)

LDT:

| LDT | COUNT(LDT) |
|--------------------|------------|
| 1 No decision | 2224 |
| 2 Sent to PC | 323 |
| 3 Approved | 194 |
| 4 Continued | 10 |
| 5 Continued to DRC | 1 |

Of the 2752 total projects, 194 applications have received the approval from LDT (Land Development & Transportation Committee)

DRC:

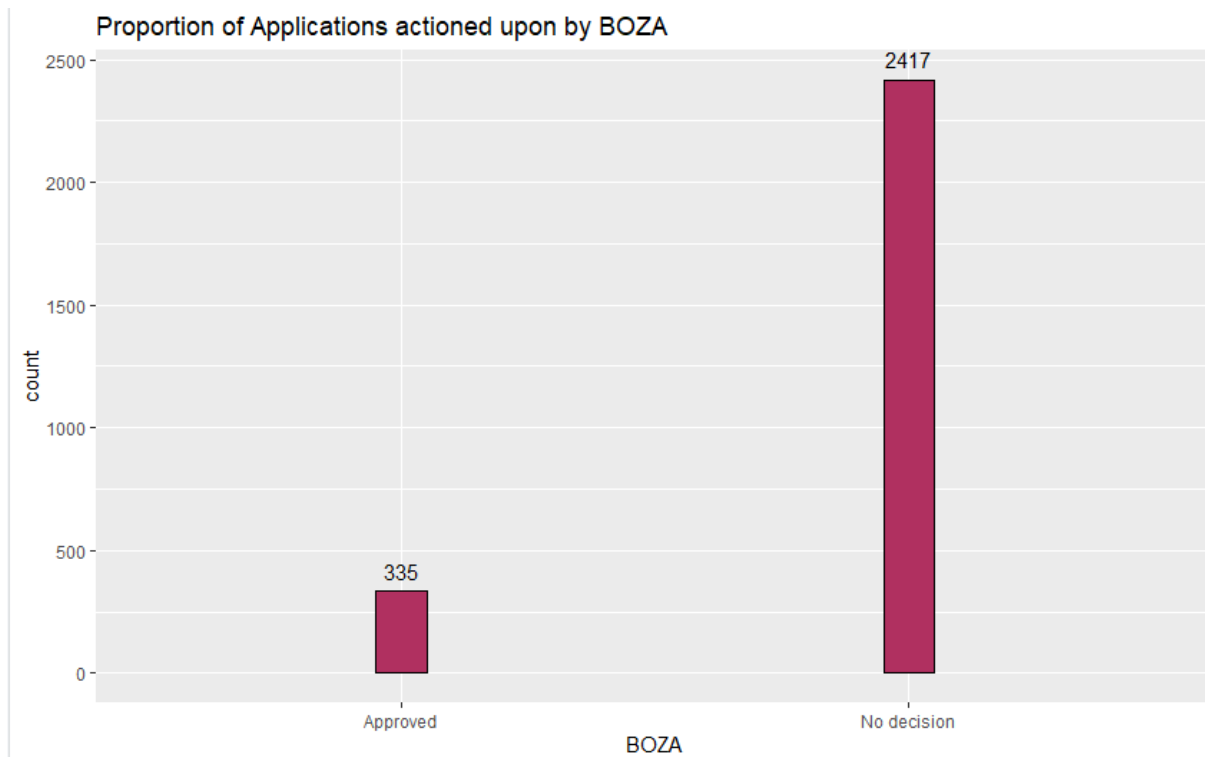
| DRC | COUNT(DRC) |
|-------------------|------------|
| 1 No decision | 2088 |
| 2 Approved | 638 |
| 3 Deferred | 2 |
| 4 Continued | 10 |
| 5 Sent to PC | 8 |
| 6 Deferred to LDT | 1 |
| 7 Approve | 5 |

Of the 2752 total projects, 638 applications have received the approval from LDT (Development Review Committee)

Approach 2: R (Visualization)

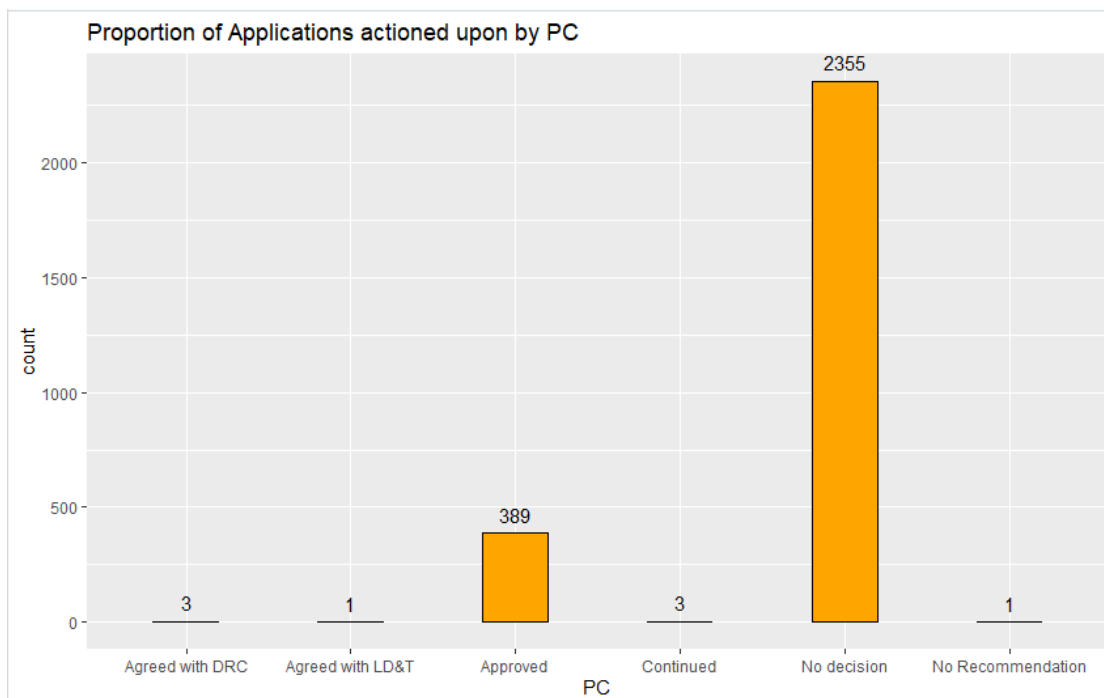
BOZA:

Plotting a bar graph to depict the proportion of projects have received the required approval of construction from BOZA:



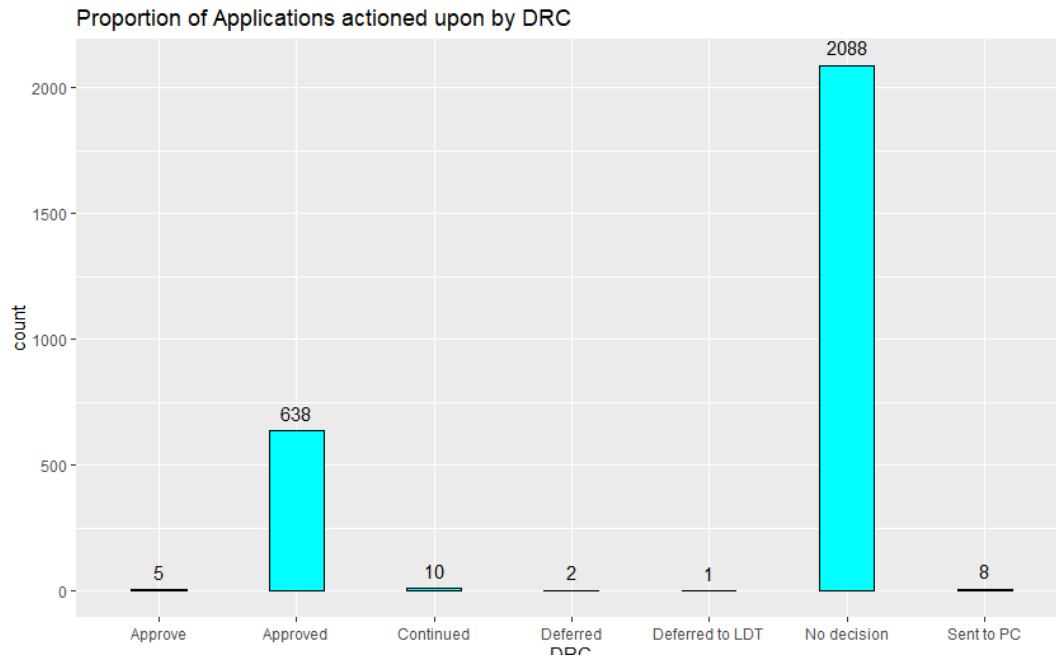
PC:

Plotting a bar graph to depict the proportion of projects have received the required approval of construction from PC:



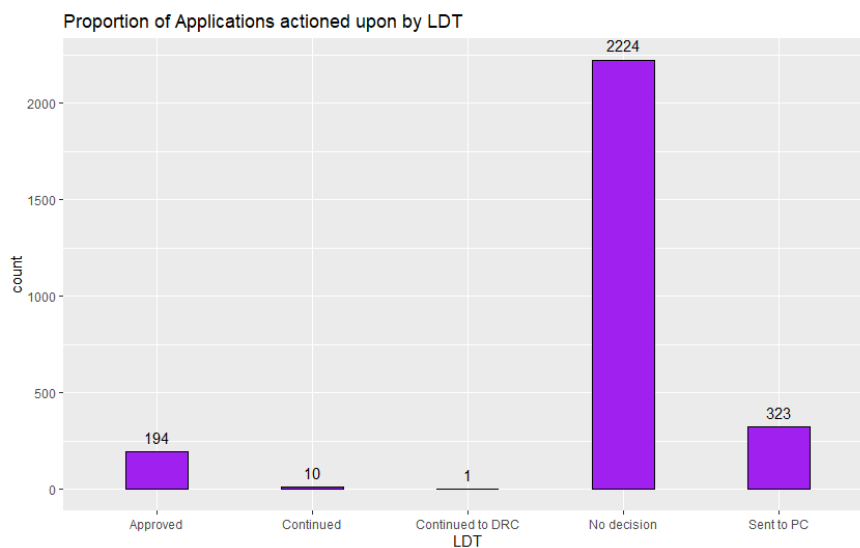
DRC:

Plotting a bar graph to depict the proportion of projects have received the required approval of construction from DRC:



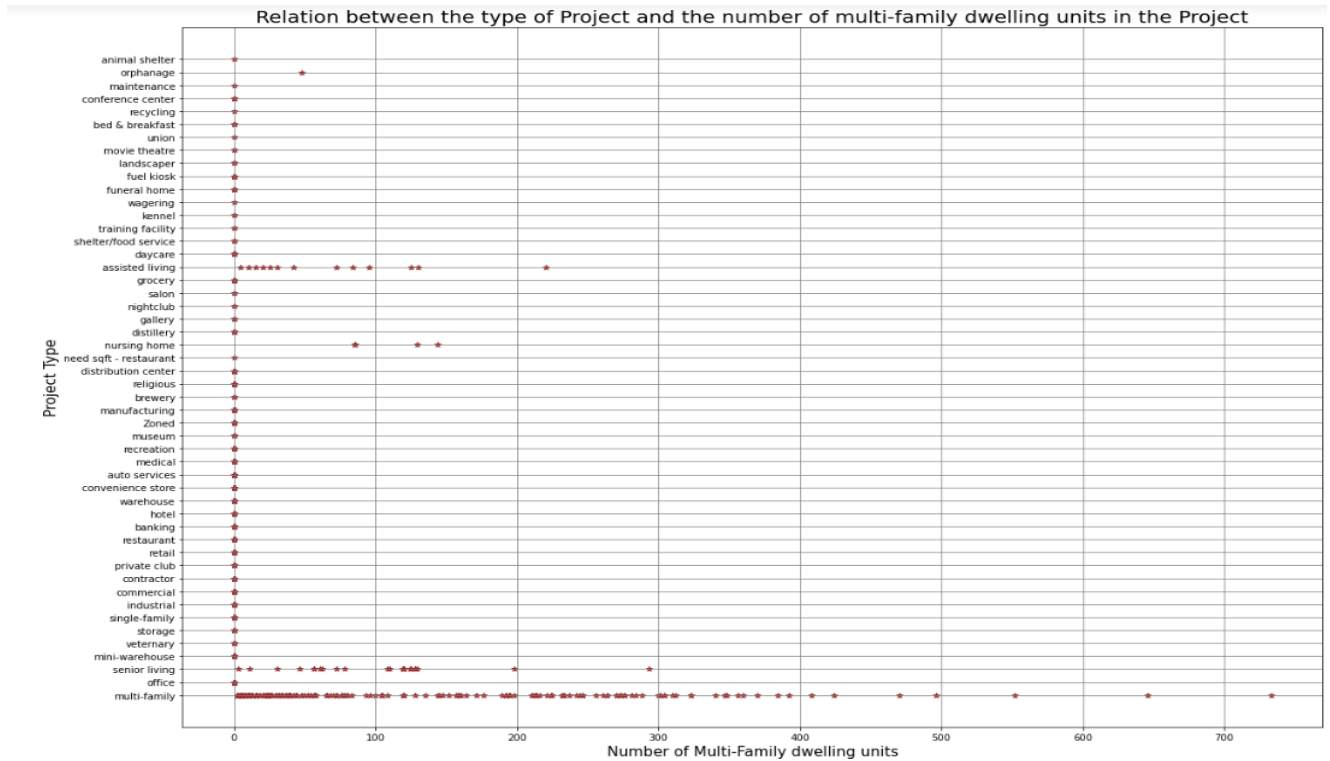
LDT:

Plotting a bar graph to depict the proportion of projects have received the required approval of construction from LDT:



Further Analysis beyond the research questions:

Relationship between Project Type and the number of multi family dwelling units in a project:



The above plot shows that the Projects that have the highest number of dwelling units belong to the 'multi-family' category, followed by the 'assisted-living' category.

Regression model to determine Project Type from the number of dwelling units:

Fitting a logistic regression model in R to determine Project Type from the number of dwelling units:

```
> development_data$ProjType <- as.factor(development_data$ProjType)
> set.seed(1000)
> split <- sample.split(development_data, splitRatio = 0.8)
> train <- subset(development_data, split == "TRUE")
> test <- subset(development_data, split == "FALSE")
```

```
> logistic_model <- glm(ProjType ~ Units, data = train, family = 'binomial')
warning messages:
1: glm.fit: algorithm did not converge
2: glm.fit: fitted probabilities numerically 0 or 1 occurred
> summary(logistic_model)
```

```
Call:
glm(formula = ProjType ~ Units, family = "binomial", data = train)
```

Deviance Residuals:

| Min | 1Q | Median | 3Q | Max |
|---------|--------|--------|--------|--------|
| -3.8953 | 0.0319 | 0.0319 | 0.0319 | 0.0319 |

Coefficients:

| | Estimate | Std. Error | z value | Pr(> z) |
|-------------|----------|------------|---------|--------------|
| (Intercept) | 7.586 | 1.000 | 7.584 | 3.34e-14 *** |
| units | 4.556 | 570.175 | 0.008 | 0.994 |

Signif. codes: 0 '***' 0.001 '**' 0.01 '*' 0.05 '.' 0.1 ' ' 1

(Dispersion parameter for binomial family taken to be 1)

Null deviance: 17.358 on 2162 degrees of freedom
Residual deviance: 17.173 on 2161 degrees of freedom
AIC: 21.173

Number of Fisher Scoring iterations: 25

As can be seen from the summary statistics of the logistic regression model, the model has a low AIC score of 21.173. This indicates that the variable 'Units' (the number of inhabitable units in a project) can be used to predict the Project Type to a certain degree.

Visualizing the Logistic Regression Model:

