EDA PROJECT

San francisco BUILDING PERMITS Dataset

Bele Rohit

Lovely Professional University

San Francisco Building Permits Dataset

# EDA Project

In Association with

Logo, company name

Description automatically generated



Submitted By : Bele Rohit

Project Name: San Francisco Building Permits Data Analysis

Section:K20RU

Registration Number: 12015904

Roll No.:K20RUB41

# Table of Contents

## Background

1. Key things about Building Permits
2. Why Permits are important?
3. Time required to get the building permit.

## Introduction

1. About the project
2. Challenges faced during the data analysis.

## Data Preparation

1. Data cleaning
2. Missing Values Analysis
3. Further Analysis with Visualizations and Insights

## Conclusions

## Reference

# Background

Building permits are a type of authorization that must be granted by a government or other regulatory body before the construction of a new or existing building can legally occur. The U.S. Census Bureau reports the finalized number of the total monthly building permits on the 18th working day of every month.

Key Things to know about the Building Permits:

* You may not need a building permit for remodelling, but you need to check with your city's building department to make sure.
* Building without a permit when one is required is unwise and can result in significant penalties.
* Building permits take into account weather-related concerns, such as earthquakes, flooding, or heavy snow.
* All electrical wiring requires a permit.
* In rare instances or when a market is especially busy, the process of receiving a permit may take longer than the valid time allowed to work on your project once the permit is issued.

Why Building Permits are Important?

* The issuance of residential building permits can be a barometer for consumer confidence and solvency.
* The U.S. Census Bureau conducts surveys of building permits for new housing units with data available for monthly, annual, and year-to-date intervals.
* The surveys cover the nation and are divided by region, state, metropolitan area, and county.
* Building permits for new housing can run the gamut from multifamily units to single-family construction.
* A general increase in building permits might indicate a need for more homes.
* A rise in building permits specifically for single-family homes can indicate that more citizens have accumulated enough finances to afford their residences.

One more question that is crucial for further proceeding is, how long does it take to get the building permit?

* The time it takes to get a permit differs greatly based on the type of permit and where you are submitting it.
* For example, you may be issued a permit on the spot if it is something fairly simple.
* On the other hand, if you submit an intricate permit in a city that is backlogged with permits, it can take much longer, sometimes even months.
* However, most non-complex permits are issued within a manner of weeks.

# INTRODUCTION

The project covered here is all about San Francisco’ s building permits. It tries to depict various parameters, and factors that affect the issuance of a building Permit. Demonstrates the various parameter by plotting the graphs with variables to visualize the data in a better and pictorial format. In one of the reports published [Trulia](https://www.trulia.com/research/elasticity-2016/) it states that “During *the last two decades San Francisco and Pittsburgh have been stingy when it comes to homebuilding and prices have risen. Contrast those metros with Las Vegas and Atlanta, some of the biggest foreclosure casualties in the housing crisis.* From this we can infer that San Francisco and Pittsburgh are facing housing scarcity.

Few of most of the questions that I’ve tried to deal which arise after the thorough analysis of the dataset.

* Can you try predicting permit issue times for various permit types? Which ones matter more?
* Can you suggest which is the best week day to visit Department of building inspections, based on this data?
* Can you conclude anything on the city's developments based on this data

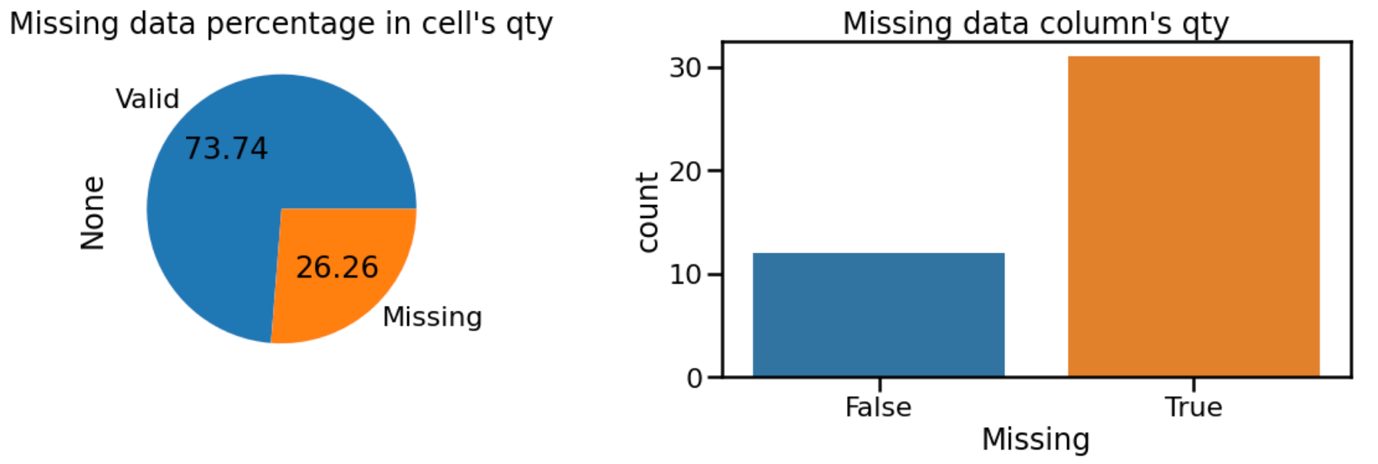
# Data Preparation

## Data Cleaning

* There are 198900 rows and 42 columns in the original dataset.

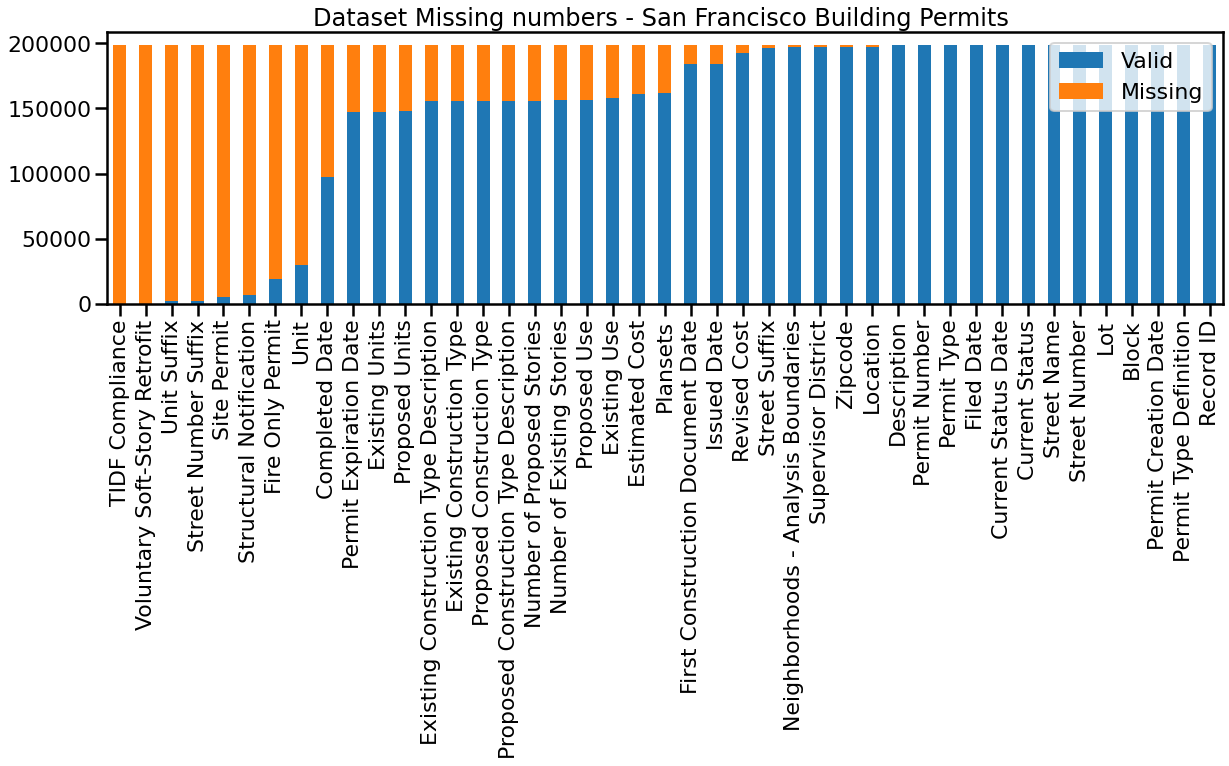
|  |  |
| --- | --- |
| Data types | Amount in Dataset |
| Datetime64 | 2 Columns |
| Float64 | 12 Columns |
| Int64 | 3 Columns |
| object | 25 Columns |

* Then instead of traversing through all the dataset I just made a list of features having more than 10k null values.
* There are 26 columns with more than 10K null values.
* There are 11 categorical columns.
* There are 6 Numerical Columns.
* Missing Data in the total data set is 26.26%
* This can be viewed pictorially as



## We can see from the above Pie chart that almost 26.26% of the total data is missing and the quantity of the valid data is 73.74%.

Now we’ll see the missing values according to a particular column in the dataset.



## Now we’ll see which column has the highest missing values in the dataset so that we can drop them easily.

The column Zip code has the highest number of missing values. This can be seen in the graph as

Chart, pie chart

Description automatically generated

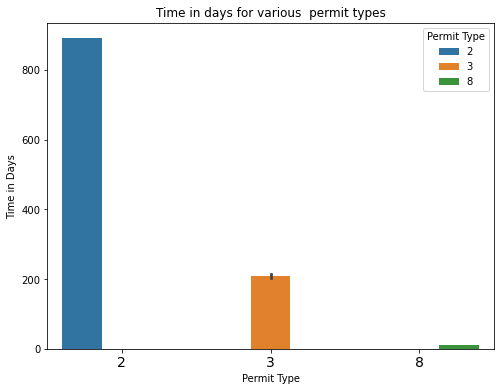
* If we were to drop columns with at least one missing value, Then, a lot of data is lost.
* Out of 42 columns only 11 are retained, 31 columns are lost.
* If we were to drop rows with at least one missing value, then, all the data would be lost.
* With the above-mentioned conditions, I’ll be dropping the columns with more than 10000 null values in it.

# Further Analysis

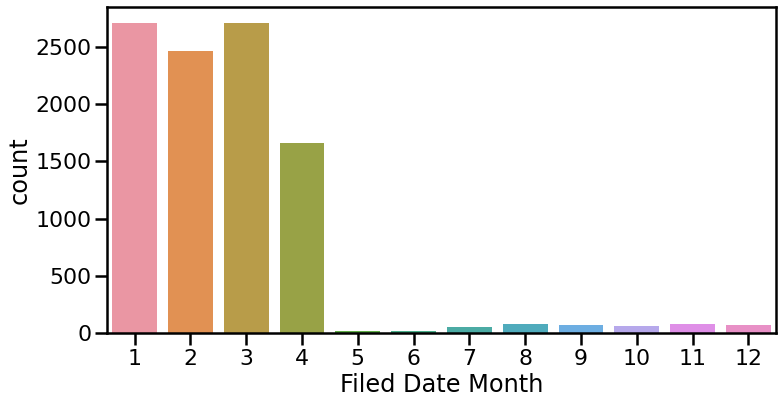
Shape

Description automatically generated

The above graph shows us that the count of days vs Time in days where time in days is calculated based on the difference of the issued date and the filed date. We can see that the graph above is highly skewed towards the left denoting that time taken is much lesser as compared to the count of records.



The above graph shows us that Permit type 2 takes the highest time to get issued by the government authority. Because Type 2 permit is the type of permit which permits several different types of work but don’t change how the building is used. Such as plumbing and interior renovation. Also permit type 8 takes the least to process.

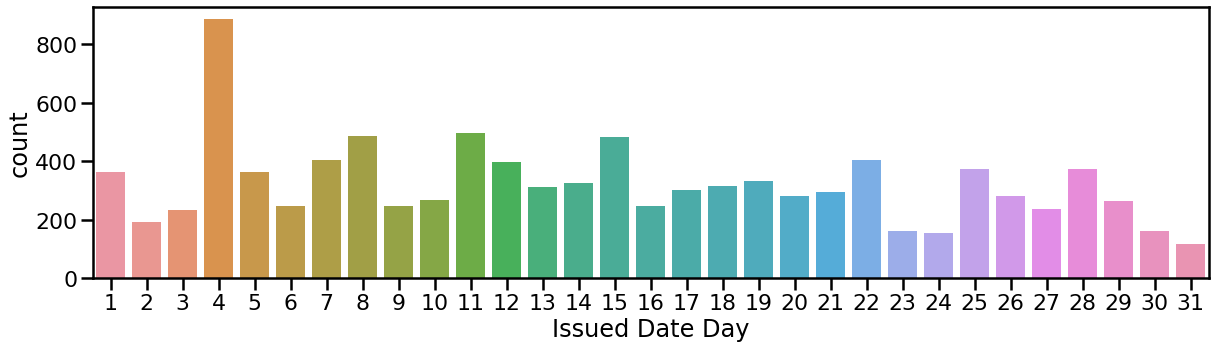


We can see from the above graph that most of the permits are filed in the first four months of the year.

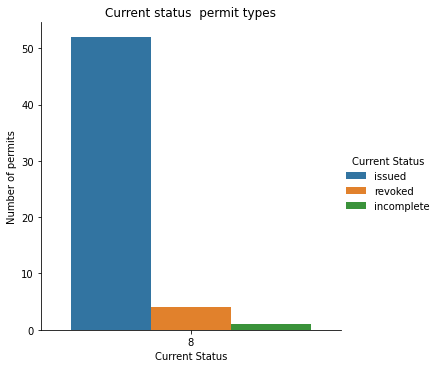
Chart

Description automatically generated

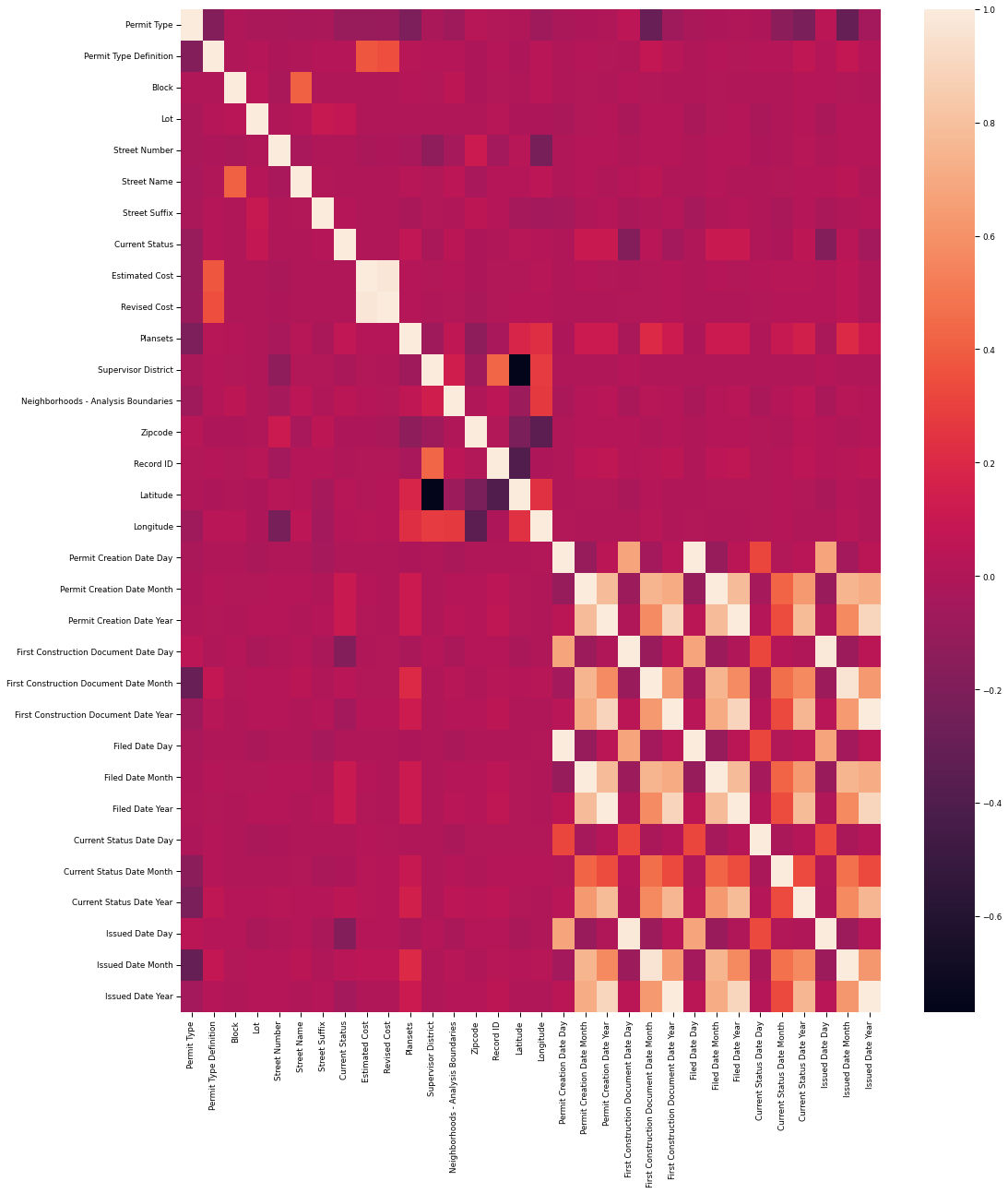
As expected, most of the permits are also issued in the first four months of the year and very few of them would of them takes more than that. This permit could of type 2 as we know that they takes longer to get issued.



The Fourth of every months seems to be the favourite date of the authorities for issuing permits.



The Bar Graph above shows that there are more permits that are issued, a few permits that are revoked and a slight number of them that are incomplete meaning that there are more chances for a permit to be issued once all required documents have been submitted.



There a decent correlation between the months of Issues Date and Filed Date features so got to look into those. Out of all these, I will be completely ignoring the features such as Street Names and Prefix along with record ID among others. This is because they don't provide enough insights or provide insights that are already too obvious. For example there is significant correlation between the Latitude / Longitude and Zip code features which is quite obvious.

Chart, waterfall chart

Description automatically generated

Graph above shows that Alterations permits are the common permits type definition in the data with other permits definitions greatly skewed.

A picture containing histogram

Description automatically generated

The most construction types permits issued belong to Existing Construction type 5.0 which is reflected across the major permit types 8 and 3, then followed by construction type 1.0 ,3.0,2.0 and least number of permits belong to construction type 4.0.

Chart, pie chart

Description automatically generated

The above graph shows the distribution according to the building type. Almost 77.2% of all data residential type of data has the highest permits.

# Conclusions

There are a key issues that we draw from the San Francisco Building Permits data:

Permit type of category 8 which involve alterations in buildings tend to matter the most and normally tend to take the least time literary in hours for the permit to be issued on average, then followed by permit type of category 3.

It's important to note that some permit types of category 2 takes the most time on average to process up to 3 years.

It can be noted that the city's most existing construction type fall in the category of type 5.0 then followed by construction type 1.0.

Most of the insights are mentioned along with graphs.

# Reference:

* [Kaggle for San Francisco Dataset](https://www.kaggle.com/datasets/aparnashastry/building-permit-applications-data)
* [Trulia.Com for Elasticity Article](https://www.trulia.com/research/elasticity-2016/)