Written Assessment - Training Program

2025-05-05

R Markdown

clear env

```
rm(list = ls())
```

Libraries

```
library(tidyverse)
```

```
## -- Attaching core tidyverse packages ------------------------ tidyverse 2.0.0 --
## v dplyr     1.1.4     v readr     2.1.5
## v forcats   1.0.0     v stringr   1.5.1
## v ggplot2   3.5.2     v tibble    3.2.1
## v lubridate 1.9.4     v tidyr     1.3.1
## v purrr     1.0.4
## -- Conflicts ------------------------------------------ tidyverse_conflicts() --
## x dplyr::filter() masks stats::filter()
## x dplyr::lag()    masks stats::lag()
## i Use the conflicted package (<http://conflicted.r-lib.org/>) to force all conflicts to become errors
```

```
library(here)
```

```
## here() starts at /home/wasif_pclab2/WK_Rprojects/RP_250505_Leaders_in_Training_Written_Assessment
```

create a new folder to save the data

```
# # Create a new folder named "Data" in your working directory
# dir.create(here("Data"), showWarnings = FALSE)
#
# # Define the url and destination file path
# url = "https://data.lacity.org/resource/9w5z-rg2h.csv"
# dest_file = here("Data/lacity_data.csv")
#
# # Download the CSV file to the new folder
# download.file(url, destfile = dest_file, mode = "wb")
```

```
#
# # Read the CSV file from the saved location
# data = read.csv(dest_file)
#
# # Preview the data
# head(data)
```

The HTTP address did not produe all the entries , so downloaded the data manually

Can not download the data from the 2nd link provided in the email

read data

```
bldg_safty_insp_data = read_csv(here("Data", "Building_and_Safety_Inspections_20250505.csv")) %>%
  # clean names
  janitor::clean_names()
```

```
## Rows: 10396028 Columns: 7
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (7): ADDRESS, PERMIT, Permit Status, Inspection Date, Inspection Type, I...
##
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
```

```
bldg_safty_insp_data %>% head(50)
```

```
## # A tibble: 50 x 7
##    address                   permit permit_status inspection_date inspection_type
##    <chr>                     <chr>  <chr>         <chr>           <chr>
##  1 10000 W SANTA MONICA BL~ 14044~ Issued        07/20/2016      Rough-Ventilat~
##  2 1000 S SANTA FE AVE      15016~ Permit Final~ 07/22/2016      Smoke Detectors
##  3 3680 N BUENA PARK DR     15014~ Issued        07/18/2016      Insulation
##  4 1001 N LINDENWOOD LANE   16042~ Permit Final~ 07/20/2016      Final
##  5 2836 S ANCHOR AVE        15016~ CofO Issued   07/18/2016      Inspection
##  6 2836 S ANCHOR AVE        15016~ CofO Issued   07/18/2016      Inspection
##  7 5489 E KEATS ST          16042~ Permit Final~ 07/18/2016      Final
##  8 4125 N PERLITA AVE #B    16016~ Issued        07/18/2016      Drywall Nailing
##  9 5744 W MANCHESTER AVE    01020~ Issued        07/22/2016      Plumbing Verif~
## 10 5924-5926 N FIGUEROA ST  16042~ Issued        07/20/2016      Rough
## # i 40 more rows
## # i 2 more variables: inspection_result <chr>, latitude_longitude <chr>
```

Count the unique number of entries in each column

```
bldg_safty_insp_data %>%
  summarise(across(everything(), ~ n_distinct(.)))
```

```
## # A tibble: 1 x 7
##   address permit permit_status inspection_date inspection_type inspection_result
##     <int>  <int>         <int>           <int>           <int>             <int>
## 1  640011 1.73e6            47            4055             185                65
## # i 1 more variable: latitude_longitude <int>
```

convert to factors if entries are below a threshold

```
bldg_safty_insp_data %>%
  mutate(across(where(~ is.character(.) && n_distinct(.) < 100), as.factor)) %>%
  head()
```

```
## # A tibble: 6 x 7
##   address permit permit_status inspection_date inspection_type inspection_result
##   <chr>   <chr>  <fct>         <chr>           <chr>           <fct>
## 1 10000 ~ 14044~ Issued        07/20/2016      Rough-Ventilat~ Partial Approval
## 2 1000 S~ 15016~ Permit Final~ 07/22/2016      Smoke Detectors Insp Cancelled
## 3 3680 N~ 15014~ Issued        07/18/2016      Insulation      Approved
## 4 1001 N~ 16042~ Permit Final~ 07/20/2016      Final           Permit Finaled
## 5 2836 S~ 15016~ CofO Issued   07/18/2016      Inspection      Permit Finaled
## 6 2836 S~ 15016~ CofO Issued   07/18/2016      Inspection      Permit Finaled
## # i 1 more variable: latitude_longitude <chr>
```

Question 1

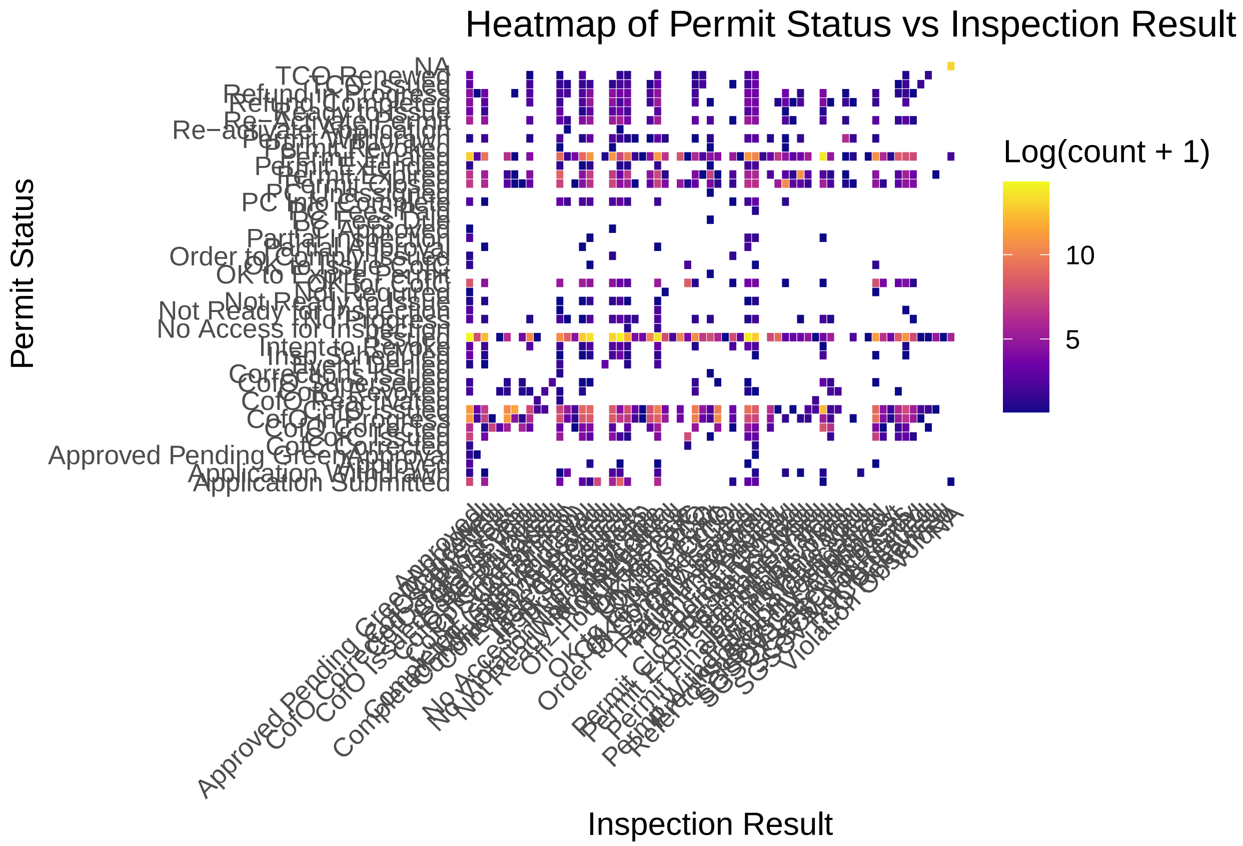
Summary Table for Permits vs Inspection results

```
summary_table = bldg_safty_insp_data %>%
  count(`permit_status`, `inspection_result`, sort = TRUE, name = "count")

# Display the table
summary_table
```

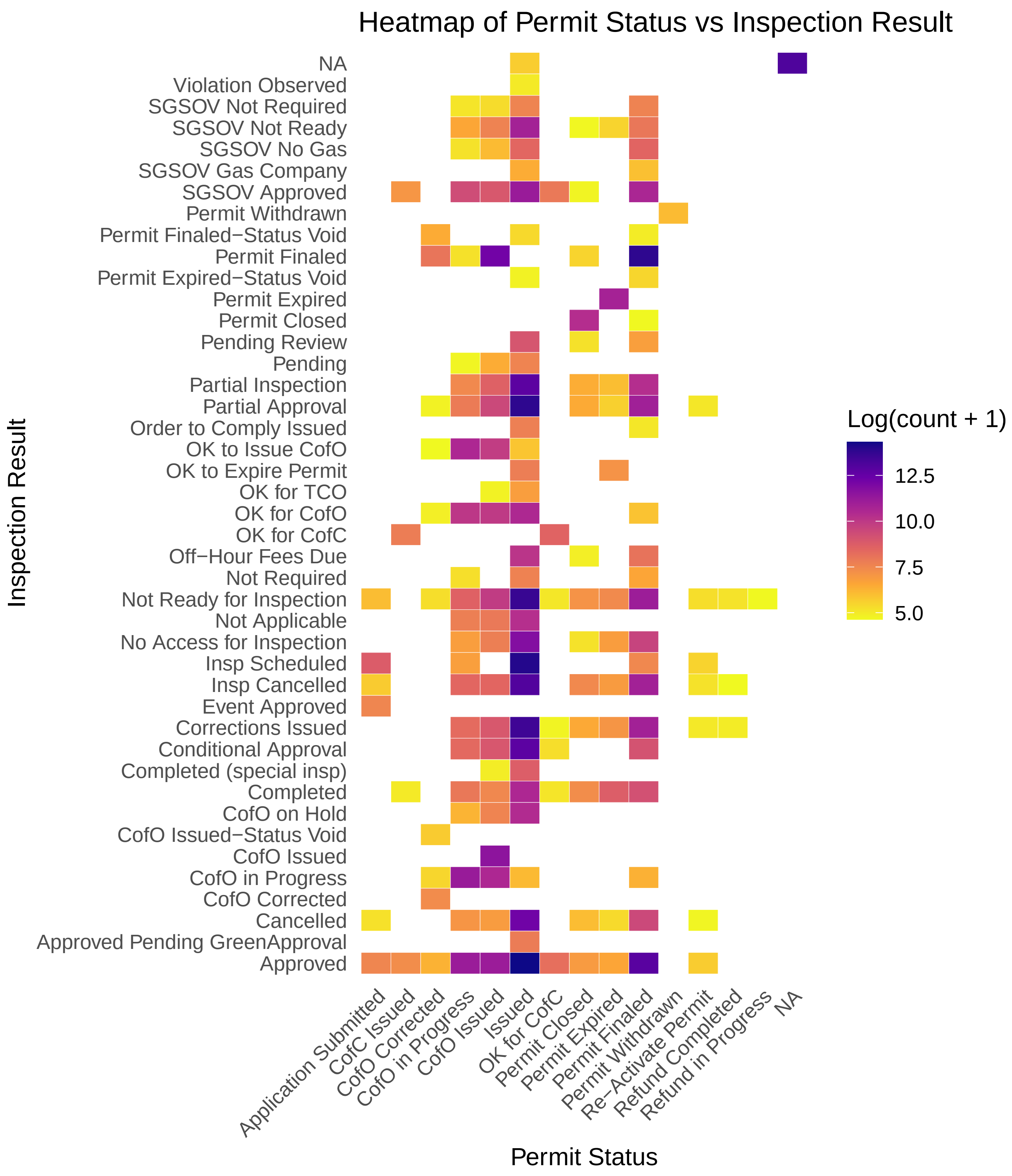
```
## # A tibble: 676 x 3
##    permit_status  inspection_result          count
##    <chr>          <chr>                      <int>
##  1 Issued         Approved                 1630424
##  2 Issued         Insp Scheduled           1232436
##  3 Permit Finaled Permit Finaled           1025570
##  4 Issued         Partial Approval         1013422
##  5 Issued         Not Ready for Inspection  838455
##  6 Issued         Corrections Issued        715577
##  7 <NA>           <NA>                      481120
##  8 Issued         Insp Cancelled            450619
##  9 Permit Finaled Approved                  374104
## 10 Issued         Partial Inspection        348725
## # i 666 more rows
```

```r
bldg_safty_insp_data %>%
  count(`permit_status`, `inspection_result`, sort = TRUE, name = "count") %>%
  mutate(count_log = log1p(count)) %>%
# ggplot heatmap
ggplot( aes(x = `inspection_result`, y = `permit_status`, fill = count_log)) +
  geom_tile(color = "white") +
  # geom_text(aes(label = count), size = 3) +
  scale_fill_viridis_c(option = "C", name = "Log(count + 1)") +
  theme_minimal(base_size = 12) +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    panel.grid = element_blank()
  ) +
  labs(
    title = "Heatmap of Permit Status vs Inspection Result",
    x = "Inspection Result",
    y = "Permit Status"
  )
```



for easy visualization and meaningful observations remove counts less than 100

```r
bldg_safty_insp_data %>%
  count(`permit_status`, `inspection_result`, sort = TRUE, name = "count") %>%
  filter(count >100) %>%
  mutate(count_log = log1p(count)) %>%
# ggplot heatmap
ggplot( aes(y = `inspection_result`, x = `permit_status`, fill = count_log)) +
  geom_tile(color = "white") +
  # geom_text(aes(label = count), size = 3) +
  scale_fill_viridis_c(option = "C", direction = -1,  name = "Log(count + 1)") +
  theme_minimal(base_size = 12) +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    panel.grid = element_blank()
  ) +
  labs(
    title = "Heatmap of Permit Status vs Inspection Result",
    y = "Inspection Result",
    x = "Permit Status"
  )
```


Heatmap of Permit Status vs Inspection Result
Inspection Result
NA
Violation Observed
SGSOV Not Required
SGSOV Not Ready
SGSOV No Gas
SGSOV Gas Company
SGSOV Approved
Permit Withdrawn
Permit Finaled−Status Void
Permit Finaled
Permit Expired−Status Void
Permit Expired
Permit Closed
Pending Review
Pending
Partial Inspection
Partial Approval
Order to Comply Issued
OK to Issue CofO
OK to Expire Permit
OK for TCO
OK for CofO
OK for CofC
Off−Hour Fees Due
Not Required
Not Ready for Inspection
Not Applicable
No Access for Inspection
Insp Scheduled
Insp Cancelled
Event Approved
Corrections Issued
Conditional Approval
Completed (special insp)
Completed
CofO on Hold
CofO Issued−Status Void
CofO Issued
CofO in Progress
CofO Corrected
Cancelled
Approved Pending GreenApproval
Approved
Application Submitted
CofC Issued
CofO Corrected
CofO in Progress
CofO Issued
Issued
OK for CofC
Permit Closed
Permit Expired
Permit Finaled
Permit Withdrawn
Re−Activate Permit
Refund Completed
Refund in Progress
NA
Permit Status
Log(count + 1)
12.5
10.0
7.5
5.0

A new link for 2nd dataset is provided

downloaded this data manually

Read permit data in R

```r
bldg_permit_data = read_csv(here("Data", "Building_Permits_20250505.csv")) %>%
  # clean names
  janitor::clean_names()
```

```
## Warning: One or more parsing issues, call ‘problems()‘ on your data frame for details,
## e.g.:
##   dat <- vroom(...)
##   problems(dat)
```

```
## Rows: 1635148 Columns: 54
## -- Column specification --------------------------------------------------------
## Delimiter: ","
## chr (41): Assessor Page, Assessor Parcel, Tract, Block, Lot, Reference # (Ol...
## dbl (12): Assessor Book, Project Number, Address Start, Address End, Zip Cod...
## lgl  (1): Event Code
##
## i Use ‘spec()‘ to retrieve the full column specification for this data.
## i Specify the column types or set ‘show_col_types = FALSE‘ to quiet this message.
```

```r
bldg_permit_data %>%  head()
```

```
## # A tibble: 6 x 54
##   assessor_book assessor_page assessor_parcel tract                 block lot
##           <dbl> <chr>         <chr>           <chr>                 <chr> <chr>
## 1          5007 001           016             TR 911                <NA>  247
## 2          5539 026           008             DAYTON HEIGHTS TRACT B     9
## 3          2384 021           048             TR 6293               <NA>  96
## 4          5535 028           001             TR 1186               <NA>  28
## 5          5432 007           005             TR 8423               <NA>  220
## 6          2118 015           008             TR 7632               <NA>  8
## # i 48 more variables: reference_number_old_permit_number <chr>,
## #   pcis_permit_number <chr>, status <chr>, status_date <chr>,
## #   permit_type <chr>, permit_sub_type <chr>, permit_category <chr>,
## #   project_number <dbl>, event_code <lgl>, initiating_office <chr>,
## #   issue_date <chr>, address_start <dbl>, address_fraction_start <chr>,
## #   address_end <dbl>, address_fraction_end <chr>, street_direction <chr>,
## #   street_name <chr>, street_suffix <chr>, suffix_direction <chr>, ...
```

To check if the two datasets have some common addresses based on address columns

```r
# Find common rows based on 'permit' and 'address'
bldg_safty_insp_permit = inner_join(bldg_safty_insp_data, bldg_permit_data,
                          by = c("address"= "applicant_address_1")
                          )
```

```
## Warning in inner_join(bldg_safty_insp_data, bldg_permit_data, by = c(address = "applicant_address_1")
## i Row 23 of ‘x‘ matches multiple rows in ‘y‘.
## i Row 194740 of ‘y‘ matches multiple rows in ‘x‘.
## i If a many-to-many relationship is expected, set ‘relationship =
##   "many-to-many"‘ to silence this warning.
```

```r
print(bldg_safty_insp_permit)
```

```
## # A tibble: 3,119,430 x 60
##    address                   permit permit_status inspection_date inspection_type
##    <chr>                     <chr>  <chr>         <chr>           <chr>
##  1 10250 W SANTA MONICA BL~ 15016~ Issued        07/19/2016      Wood Frame
##  2 9045 S LINCOLN BLVD      15046~ Permit Final~ 07/20/2016      Final
##  3 9045 S LINCOLN BLVD      15046~ Permit Final~ 07/20/2016      Final
##  4 9045 S LINCOLN BLVD      15046~ Permit Final~ 07/20/2016      Final
##  5 4205 W 63RD ST           16014~ Issued        07/19/2016      Footing/Founda~
##  6 1150 W 25TH ST           16016~ Issued        07/22/2016      Final
##  7 1318 E 7TH ST            16041~ Issued        07/18/2016      Rough
##  8 1318 E 7TH ST            16041~ Issued        07/18/2016      Rough
##  9 1318 E 7TH ST            16041~ Issued        07/18/2016      Rough
## 10 1318 E 7TH ST            16041~ Issued        07/18/2016      Rough
## # i 3,119,420 more rows
## # i 55 more variables: inspection_result <chr>, latitude_longitude.x <chr>,
## #   assessor_book <dbl>, assessor_page <chr>, assessor_parcel <chr>,
## #   tract <chr>, block <chr>, lot <chr>,
## #   reference_number_old_permit_number <chr>, pcis_permit_number <chr>,
## #   status <chr>, status_date <chr>, permit_type <chr>, permit_sub_type <chr>,
## #   permit_category <chr>, project_number <dbl>, event_code <lgl>, ...
```

Some address have more than 2 entries and thats why we are getting more than 1 row for the same address there are more rows than the permit dataset

We can now select the columns we are interested in and save the data

```r
bldg_safty_insp_permit_2 = bldg_safty_insp_permit %>%
  select(
    address,
    permit_status,
    inspection_type,
    inspection_result,
    status,
    permit_type,
    contractor_city,
    contractor_state,
    applicant_address_3,
    zone
    ) %>%
  # remove duplicates
  distinct()
bldg_safty_insp_permit_2 %>%  dim()
```

```
## [1] 681183     10
```

```
bldg_safty_insp_permit_2 %>% head(15)
```

```
## # A tibble: 15 x 10
##    address    permit_status inspection_type inspection_result status permit_type
##    <chr>      <chr>         <chr>           <chr>             <chr>  <chr>
##  1 10250 W S~ Issued        Wood Frame      Partial Approval  Permi~ Bldg-Alter~
##  2 9045 S LI~ Permit Final~ Final           Permit Finaled    Permi~ Bldg-Alter~
##  3 9045 S LI~ Permit Final~ Final           Permit Finaled    CofO ~ Bldg-Alter~
##  4 4205 W 63~ Issued        Footing/Founda~ Approved          Issued Bldg-Addit~
##  5 1150 W 25~ Issued        Final           Insp Scheduled    CofO ~ Bldg-Addit~
##  6 1318 E 7T~ Issued        Rough           Partial Inspecti~ Permi~ Nonbldg-New
##  7 1318 E 7T~ Issued        Rough           Partial Inspecti~ Permi~ Bldg-Alter~
##  8 1318 E 7T~ Issued        Rough           Partial Inspecti~ Permi~ Bldg-Alter~
##  9 1318 E 7T~ Issued        Rough           Partial Inspecti~ CofO ~ Bldg-Alter~
## 10 1318 E 7T~ Issued        Rough           Partial Inspecti~ Issued Bldg-New
## 11 1318 E 7T~ Issued        Rough           Partial Inspecti~ Permi~ Bldg-Alter~
## 12 1318 E 7T~ Issued        Rough           Partial Inspecti~ Issued Bldg-Alter~
## 13 3911 S FI~ Issued        Footing/Founda~ Approved          Permi~ Electrical
## 14 21650 W O~ Issued        Drywall Nailing Approved          CofO ~ Bldg-New
## 15 21650 W O~ Issued        Drywall Nailing Approved          Permi~ Bldg-Alter~
## # i 4 more variables: contractor_city <chr>, contractor_state <chr>,
## #   applicant_address_3 <chr>, zone <chr>
```

I notices that LOS ANGELES had multiple entries eg LOS ANGELES, CA, LOS ANGELES ,CA, Los Angeles, L.A., CA, etc.

The code below is to unify it

```
bldg_safty_insp_permit_2 = bldg_safty_insp_permit_2 %>%
  mutate(
    applicant_address_3 = str_to_upper(applicant_address_3),                  # Make all uppercase
    applicant_address_3 = str_replace_all(applicant_address_3, "[[:punct:]]", ""), # Remove punctuation
    applicant_address_3 = str_squish(applicant_address_3),                    # Remove extra spaces
    applicant_address_3 = case_when(
      str_detect(applicant_address_3, "LOS ANGELES") ~ "LOS ANGELES, CA",
      str_detect(applicant_address_3, "L A") ~ "LOS ANGELES, CA",
      str_detect(applicant_address_3, "L\\.A") ~ "LOS ANGELES, CA",
      str_detect(applicant_address_3, "LA CA") ~ "LOS ANGELES, CA",
      str_detect(applicant_address_3, "LACA ") ~ "LOS ANGELES, CA",
      is.na(applicant_address_3) ~ "UNKNOWN",
      TRUE ~ applicant_address_3
    )
  )
```

To get an idea of number of inspection by geography we can use applicant_address_3 which is the city

and we can also use zone

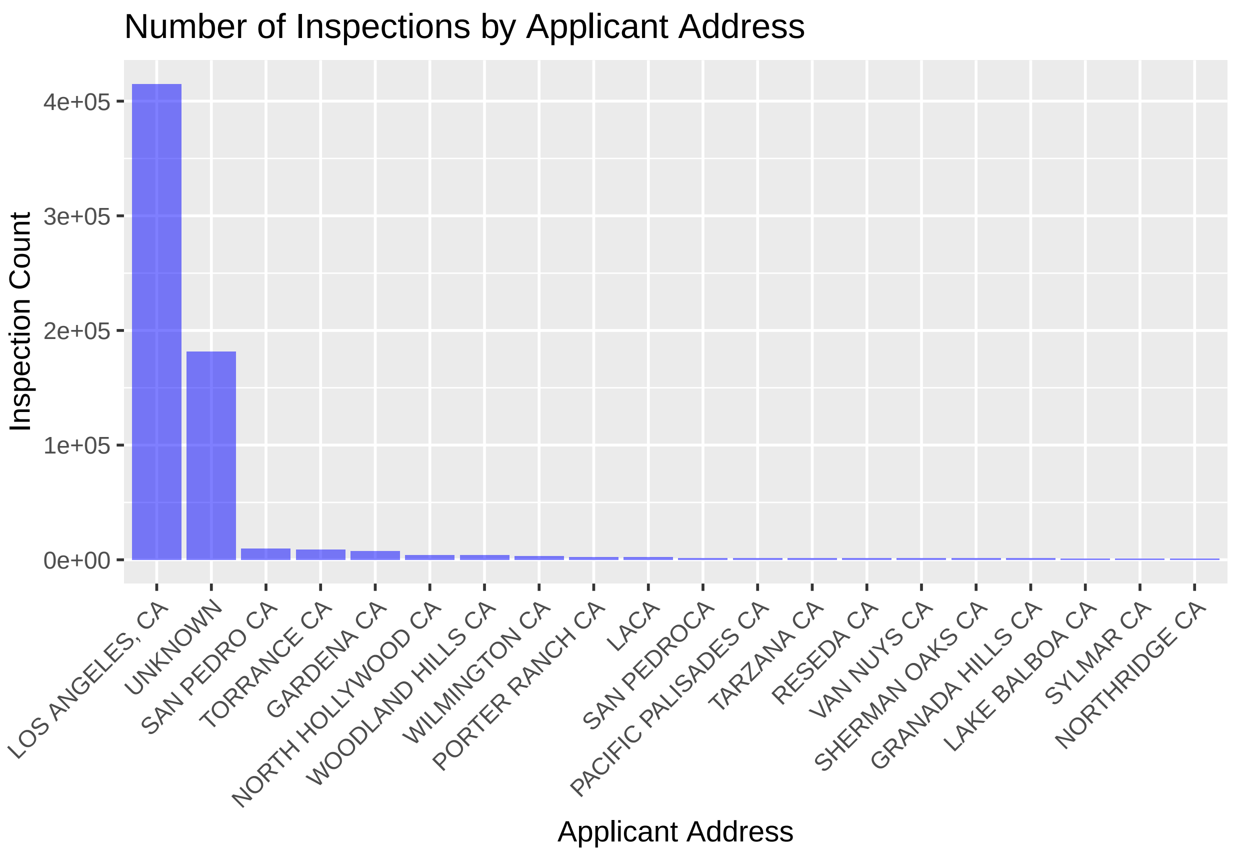
```
# Count inspections by applicant_address_3
inspection_freq = bldg_safty_insp_permit_2 %>%
  count(applicant_address_3, name = "inspection_count") %>%
  arrange(desc(inspection_count))

print(inspection_freq)
```

```
## # A tibble: 375 x 2
##    applicant_address_3 inspection_count
##    <chr>                          <int>
##  1 LOS ANGELES, CA               414938
##  2 UNKNOWN                       181791
##  3 SAN PEDRO CA                    9595
##  4 TORRANCE CA                     8952
##  5 GARDENA CA                      7561
##  6 NORTH HOLLYWOOD CA              3981
##  7 WOODLAND HILLS CA               3960
##  8 WILMINGTON CA                   3071
##  9 PORTER RANCH CA                 2441
## 10 LACA                            2304
## # i 365 more rows
```

PLot for inspection frequency for top 20

```
inspection_freq %>%
  # filter top 20 rows
  slice_head(n = 20) %>%
  # Make names as factors to avoid alphabetical ordering
  mutate(applicant_address_3 = factor(applicant_address_3, levels = inspection_freq$applicant_address_3)
  ggplot(aes(x = applicant_address_3,  y = inspection_count)) +
  geom_col(fill = "blue", alpha = 0.5) +
  labs(title = "Number of Inspections by Applicant Address",
       x = "Applicant Address", y = "Inspection Count") +
  # rotate x-axis labels
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
```



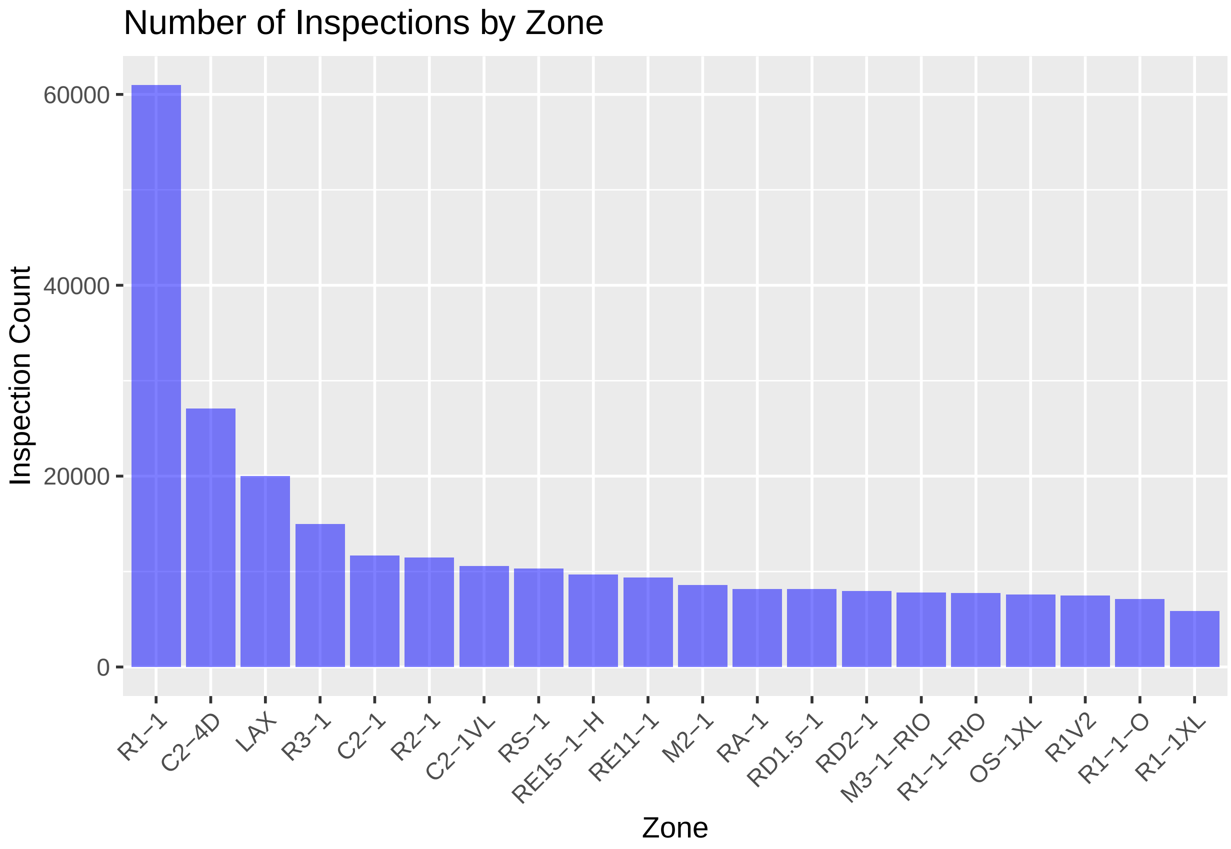
## Lets try the same thing by the zone

```
# Count inspections by zone
inspection_freq_zone = bldg_safty_insp_permit_2 %>%
  count(zone, name = "inspection_count") %>%
  arrange(desc(inspection_count))
print(inspection_freq_zone)
```

```
## # A tibble: 1,736 x 2
##    zone      inspection_count
##    <chr>                <int>
##  1 R1-1                 60950
##  2 C2-4D                27057
##  3 LAX                  19983
##  4 R3-1                 14967
##  5 C2-1                 11664
##  6 R2-1                 11475
##  7 C2-1VL               10532
##  8 RS-1                 10286
##  9 RE15-1-H              9643
## 10 RE11-1                9341
## # i 1,726 more rows
```

PLot for inspection frequency for top 20

```
inspection_freq_zone_plot =
  inspection_freq_zone %>%
  # filter top 20 rows
  slice_head(n = 20) %>%
  # Make names as factors to avoid alphabetical ordering
  mutate(zone = factor(zone, levels = inspection_freq_zone$zone)) %>%
  ggplot(aes(x = zone,  y = inspection_count)) +
  geom_col(fill = "blue", alpha = 0.5) +
  labs(title = "Number of Inspections by Zone",
       x = "Zone", y = "Inspection Count") +
  # rotate x-axis labels
  theme(axis.text.x = element_text(angle = 45, hjust = 1))
print(inspection_freq_zone_plot)
```



Inspection result by applicant address, considered as geography

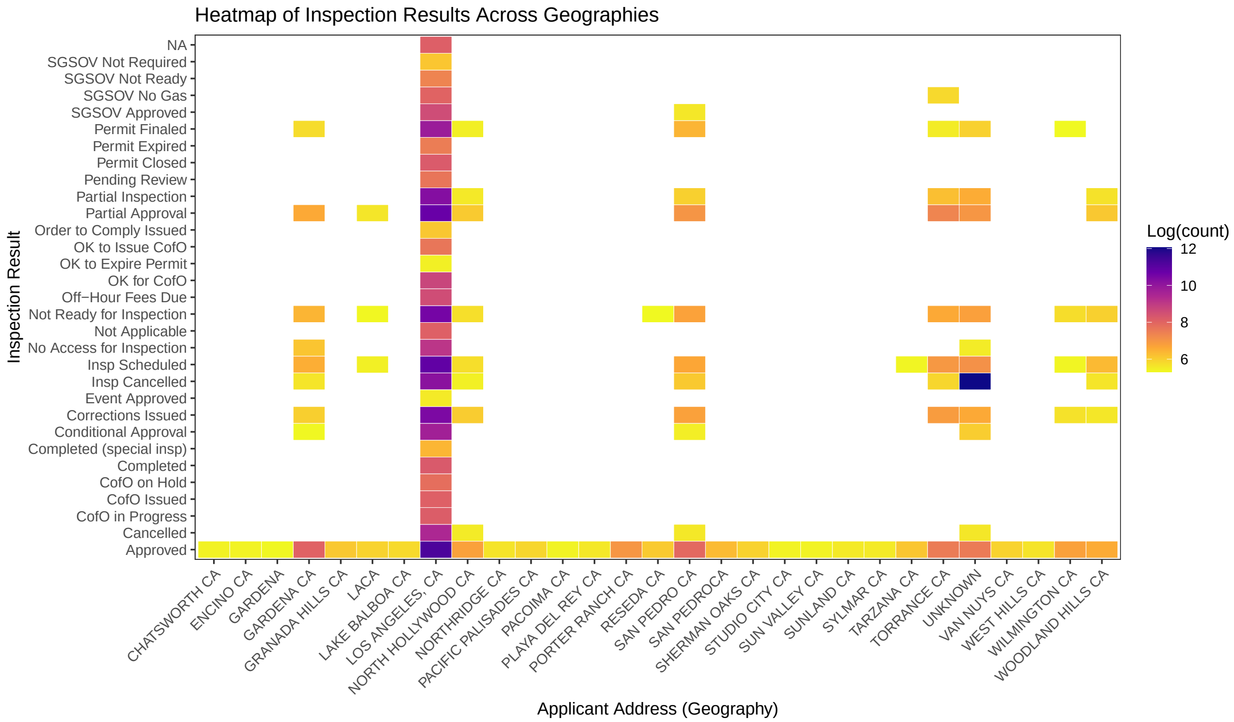
```
inspection_result_tbl = bldg_safty_insp_permit_2 %>%
  count(applicant_address_3, inspection_result, name = "count") %>%
  # order by inspection result
  arrange(desc(count))
```

```r
# View the table
print(inspection_result_tbl)
```

```
## # A tibble: 3,060 x 3
##    applicant_address_3 inspection_result         count
##    <chr>               <chr>                     <int>
##  1 UNKNOWN             Insp Cancelled           172637
##  2 LOS ANGELES, CA     Approved                  77118
##  3 LOS ANGELES, CA     Insp Scheduled            53058
##  4 LOS ANGELES, CA     Partial Approval          46813
##  5 LOS ANGELES, CA     Not Ready for Inspection  37748
##  6 LOS ANGELES, CA     Corrections Issued        33544
##  7 LOS ANGELES, CA     Partial Inspection        29241
##  8 LOS ANGELES, CA     Insp Cancelled            25926
##  9 LOS ANGELES, CA     Permit Finaled            18998
## 10 LOS ANGELES, CA     Conditional Approval      16366
## # i 3,050 more rows
```

for easy visualization and meaningful observations remove counts less than 200

```r
inspection_result_tbl %>%
  filter(count > 200) %>%
  mutate(
    count_log = log1p(count),
    applicant_address_3 = ifelse(is.na(applicant_address_3), "Unknown", applicant_address_3)
  ) %>%
  ggplot(aes(x = applicant_address_3, y = inspection_result, fill = count_log)) +
  geom_tile(color = "white") +
  scale_fill_viridis_c(option = "C", direction = -1, name = "Log(count)") +
  theme_bw(base_size = 12) +
  theme(
    axis.text.x = element_text(angle = 45, hjust = 1),
    panel.grid = element_blank()
  ) +
  labs(
    title = "Heatmap of Inspection Results Across Geographies",
    x = "Applicant Address (Geography)",
    y = "Inspection Result"
  )
```



## To answer question 3 we will use only Los Angeles data

```
bldg_safty_insp_permit_2_LA = bldg_safty_insp_permit_2 %>%
  filter(applicant_address_3 == "LOS ANGELES, CA") %>%
  # remove duplicates
  distinct()
bldg_safty_insp_permit_2_LA %>%  dim()
```

```
## [1] 399385     10
```

```
bldg_safty_insp_permit_2_LA %>% head(15)
```

```
## # A tibble: 15 x 10
##    address    permit_status inspection_type inspection_result status permit_type
##    <chr>      <chr>         <chr>           <chr>             <chr>  <chr>
##  1 10250 W S~ Issued        Wood Frame      Partial Approval  Permi~ Bldg-Alter~
##  2 9045 S LI~ Permit Final~ Final           Permit Finaled    Permi~ Bldg-Alter~
##  3 9045 S LI~ Permit Final~ Final           Permit Finaled    CofO ~ Bldg-Alter~
##  4 4205 W 63~ Issued        Footing/Founda~ Approved          Issued Bldg-Addit~
##  5 1318 E 7T~ Issued        Rough           Partial Inspecti~ Permi~ Nonbldg-New
##  6 1318 E 7T~ Issued        Rough           Partial Inspecti~ Permi~ Bldg-Alter~
##  7 1318 E 7T~ Issued        Rough           Partial Inspecti~ Permi~ Bldg-Alter~
##  8 1318 E 7T~ Issued        Rough           Partial Inspecti~ CofO ~ Bldg-Alter~
##  9 1318 E 7T~ Issued        Rough           Partial Inspecti~ Issued Bldg-New
## 10 1318 E 7T~ Issued        Rough           Partial Inspecti~ Issued Bldg-Alter~
## 11 3911 S FI~ Issued        Footing/Founda~ Approved          Permi~ Electrical
## 12 225 E 31S~ Issued        Green Building~ Not Ready for In~ Issued Bldg-Addit~
## 13 1026 S BR~ Issued        Rough           Insp Scheduled    Permi~ Bldg-Demol~
## 14 1026 S BR~ Issued        Rough           Insp Scheduled    Permi~ Bldg-Alter~
## 15 1026 S BR~ Issued        Rough           Insp Scheduled    Permi~ Bldg-Demol~
## # i 4 more variables: contractor_city <chr>, contractor_state <chr>,
## #   applicant_address_3 <chr>, zone <chr>
```

Check if the naming of contractor city is unified

```
bldg_safty_insp_permit_2_LA %>%
  count(contractor_city, name = "count") %>%
  arrange(desc(count)) %>%
  print(n=100)
```

```
## # A tibble: 344 x 2
##     contractor_city   count
##     <chr>             <int>
##   1 <NA>             103324
##   2 LOS ANGELES       93476
##   3 ANAHEIM            8256
##   4 SAN FRANCISCO      7391
##   5 IRVINE             6318
##   6 WOODLAND HILLS     5197
##   7 GLENDALE           4922
##   8 MARTINEZ           4671
##   9 MORRISTOWN         4368
##  10 SANTA FE SPRINGS   4251
##  11 ORANGE             4203
##  12 PASADENA           3953
##  13 BEVERLY HILLS      3802
##  14 RANCHO DOMINGUEZ   3714
##  15 SANTA MONICA       3597
##  16 LONG BEACH         3113
##  17 FULLERTON          2981
##  18 CALABASAS          2816
##  19 VAN NUYS           2797
##  20 NORWALK            2599
##  21 BURBANK            2467
##  22 SAN DIMAS          2456
##  23 TORRANCE           2394
##  24 SIMI VALLEY        2357
##  25 SHERMAN OAKS       2318
##  26 SYLMAR             2209
##  27 CHATSWORTH         2071
##  28 NEWPORT BEACH      1958
##  29 CARSON             1956
##  30 ENCINO             1926
##  31 THOUSAND OAKS      1917
##  32 SAN DIEGO          1888
##  33 HUNTINGTON BEACH   1818
##  34 LOMITA             1736
##  35 SUN VALLEY         1686
##  36 VALENCIA           1677
##  37 SAN JOSE           1654
##  38 SANTA CLARITA      1570
##  39 MONTEREY PARK      1567
##  40 SANTA ANA          1513
##  41 SAN FERNANDO       1500
##  42 BREA               1462
##  43 BETHESDA           1440
##  44 GARDEN GROVE       1368
##  45 NORTH HOLLYWOOD    1365
##  46 CORONA             1260
##  47 TARZANA            1251
##  48 MONROVIA           1233
##  49 ALAMEDA            1205
##  50 GARDENA            1169
##  51 HOLLYWOOD          1168
##  52 CONCORD            1164
##  53 MONTROSE           1134
##  54 REDONDO BEACH      1096
##  55 NEW YORK           1090
##  56 COVINA             1063
##  57 WEST HILLS         1044
##  58 YORBA LINDA        1038
##  59 RIVERSIDE          1008
##  60 LAGUNA BEACH        963
##  61 FOUNTAIN VALLEY     954
##  62 STUDIO CITY         943
##  63 CITY OF INDUSTRY    941
##  64 PLACENTIA           941
##  65 CULVER CITY         929
##  66 WESTLAKE VILLAGE    929
##  67 HOUSTON             925
##  68 PARAMOUNT           838
##  69 TUSTIN              838
##  70 TUJUNGA             812
##  71 ANAHEIM HILLS       769
##  72 SIGNAL HILL         764
##  73 AZUSA               743
##  74 GREELEY             742
##  75 CANOGA PARK         728
##  76 SCOTTSDALE          725
##  77 SACRAMENTO          721
##  78 ONTARIO             711
##  79 COSTA MESA          710
##  80 PARSIPPANY          696
##  81 SOUTH GATE          682
##  82 CHINO HILLS         673
##  83 VENTURA             667
##  84 FONTANA             640
##  85 INGLEWOOD           613
##  86 LOS ALAMITOS        612
##  87 BUENA PARK          607
##  88 IRWINDALE           604
##  89 WEST COVINA         589
##  90 MISSION VIEJO       579
##  91 CARLSBAD            559
##  92 NORTHRIDGE          554
##  93 CLAREMONT          540
##  94 BOSTON             535
##  95 NEWBURY PARK       535
##  96 ALTADENA           529
##  97 ARCADIA            528
##  98 LA VERNE           513
##  99 CHINO              511
## 100 RANCHO CUCAMONGA   500
## # i 244 more rows
```

Yes mostly it is

now convert the contractor data to 0 (not from Los Angeles) and 1 (from Los Angeles)

```
bldg_safty_insp_permit_2_LA_2 = bldg_safty_insp_permit_2_LA %>%
  mutate(
    contractor_city_binary = case_when(
      str_detect(contractor_city, "LOS ANGELES") ~ 1,
      TRUE ~ 0
    )
  )

bldg_safty_insp_permit_2_LA_2 %>% head()
```

```
## # A tibble: 6 x 11
##   address     permit_status inspection_type inspection_result status permit_type
##   <chr>       <chr>         <chr>           <chr>             <chr>  <chr>
## 1 10250 W SA~ Issued        Wood Frame      Partial Approval  Permi~ Bldg-Alter~
## 2 9045 S LIN~ Permit Final~ Final           Permit Finaled    Permi~ Bldg-Alter~
## 3 9045 S LIN~ Permit Final~ Final           Permit Finaled    CofO ~ Bldg-Alter~
## 4 4205 W 63R~ Issued        Footing/Founda~ Approved          Issued Bldg-Addit~
## 5 1318 E 7TH~ Issued        Rough           Partial Inspecti~ Permi~ Nonbldg-New
## 6 1318 E 7TH~ Issued        Rough           Partial Inspecti~ Permi~ Bldg-Alter~
## # i 5 more variables: contractor_city <chr>, contractor_state <chr>,
## #   applicant_address_3 <chr>, zone <chr>, contractor_city_binary <dbl>
```

```
# Check if the contractor city is converted to 0 and 1
table(bldg_safty_insp_permit_2_LA_2$contractor_city) %>% head()
```

```
##
## \\LAKE FOREST    \\PASADENA  5BELL CANYON         ACTON         AGOURA
##            19            54            12           295             78
##   AGOURA HILLS
##            429
```

Now check the types of inspection outcome

```
inspection_result_count =
bldg_safty_insp_permit_2_LA_2 %>%
  count(inspection_result, name = "count") %>%
  arrange(desc(count)) %>%
  filter (count > 1000) %>%
  print(n=100)
```

```
## # A tibble: 26 x 2
##    inspection_result        count
##    <chr>                    <int>
##  1 Approved                 74767
##  2 Insp Scheduled           51164
##  3 Partial Approval         44889
##  4 Not Ready for Inspection 36247
##  5 Corrections Issued       32240
##  6 Partial Inspection       28008
##  7 Insp Cancelled           24939
##  8 Permit Finaled           18331
##  9 Conditional Approval     15676
## 10 Cancelled                12295
## 11 No Access for Inspection  8721
## 12 OK for CofO               6124
## 13 SGSOV Approved            5143
## 14 Off-Hour Fees Due         5045
## 15 Completed                 3709
## 16 Permit Closed             3608
## 17 CofO in Progress          3442
## 18 <NA>                      3297
## 19 CofO Issued               3222
## 20 Not Applicable            3160
## 21 SGSOV No Gas              2910
## 22 CofO on Hold              2386
## 23 OK to Issue CofO          2050
## 24 Pending Review            2032
## 25 Permit Expired            1806
## 26 SGSOV Not Ready           1555
```

Now we will select them and convert them to binary based on success 1 or no success 0

```
bldg_safty_insp_permit_2_LA_3 = bldg_safty_insp_permit_2_LA_2 %>%
  filter(inspection_result %in% inspection_result_count$inspection_result) %>%
           # remove duplicates
           distinct()
bldg_safty_insp_permit_2_LA_3 %>% dim()
```

```
## [1] 396766     11
```

```
bldg_safty_insp_permit_2_LA_3 %>% head(15)
```

```
## # A tibble: 15 x 11
##    address    permit_status inspection_type inspection_result status permit_type
##    <chr>      <chr>         <chr>           <chr>             <chr>  <chr>
##  1 10250 W S~ Issued        Wood Frame      Partial Approval  Permi~ Bldg-Alter~
##  2 9045 S LI~ Permit Final~ Final           Permit Finaled    Permi~ Bldg-Alter~
##  3 9045 S LI~ Permit Final~ Final           Permit Finaled    CofO ~ Bldg-Alter~
##  4 4205 W 63~ Issued        Footing/Founda~ Approved          Issued Bldg-Addit~
##  5 1318 E 7T~ Issued        Rough           Partial Inspecti~ Permi~ Nonbldg-New
##  6 1318 E 7T~ Issued        Rough           Partial Inspecti~ Permi~ Bldg-Alter~
##  7 1318 E 7T~ Issued        Rough           Partial Inspecti~ Permi~ Bldg-Alter~
##  8 1318 E 7T~ Issued        Rough           Partial Inspecti~ CofO ~ Bldg-Alter~
##  9 1318 E 7T~ Issued        Rough           Partial Inspecti~ Issued Bldg-New
## 10 1318 E 7T~ Issued        Rough           Partial Inspecti~ Issued Bldg-Alter~
## 11 3911 S FI~ Issued        Footing/Founda~ Approved          Permi~ Electrical
## 12 225 E 31S~ Issued        Green Building~ Not Ready for In~ Issued Bldg-Addit~
## 13 1026 S BR~ Issued        Rough           Insp Scheduled    Permi~ Bldg-Demol~
## 14 1026 S BR~ Issued        Rough           Insp Scheduled    Permi~ Bldg-Alter~
## 15 1026 S BR~ Issued        Rough           Insp Scheduled    Permi~ Bldg-Demol~
## # i 5 more variables: contractor_city <chr>, contractor_state <chr>,
## #   applicant_address_3 <chr>, zone <chr>, contractor_city_binary <dbl>
```

```
## Now we will convert the inspection result to binary
bldg_safty_insp_permit_2_LA_3_insp_binary = bldg_safty_insp_permit_2_LA_3 %>%
  mutate(
    inspection_binary = case_when(
      inspection_result %in% c("Approved", "Permit Finaled", "CofO Issued",
                               "OK for CofO", "OK to Issue CofO", "Completed",
                               "SGSOV Approved") ~ 1,
      inspection_result %in% c("Insp Scheduled", "Partial Approval",
                               "Not Ready for Inspection", "Corrections Issued",
                               "Partial Inspection", "Insp Cancelled",
                               "Conditional Approval", "Cancelled",
                               "No Access for Inspection", "SGSOV No Gas",
                               "CofO in Progress", "CofO on Hold",
                               "Off-Hour Fees Due", "Pending Review") ~ 0,
      TRUE ~ NA_real_  # Handles NA or unmatched values
    )
  )
bldg_safty_insp_permit_2_LA_3_insp_binary %>% dim()
```

```
## [1] 396766     12
```

```
bldg_safty_insp_permit_2_LA_3_insp_binary %>% head(15)
```

```
## # A tibble: 15 x 12
##    address    permit_status inspection_type inspection_result status permit_type
##    <chr>      <chr>         <chr>           <chr>             <chr>  <chr>
##  1 10250 W S~ Issued        Wood Frame      Partial Approval  Permi~ Bldg-Alter~
##  2 9045 S LI~ Permit Final~ Final           Permit Finaled    Permi~ Bldg-Alter~
##  3 9045 S LI~ Permit Final~ Final           Permit Finaled    CofO ~ Bldg-Alter~
##  4 4205 W 63~ Issued        Footing/Founda~ Approved          Issued Bldg-Addit~
##  5 1318 E 7T~ Issued        Rough           Partial Inspecti~ Permi~ Nonbldg-New
##  6 1318 E 7T~ Issued        Rough           Partial Inspecti~ Permi~ Bldg-Alter~
##  7 1318 E 7T~ Issued        Rough           Partial Inspecti~ Permi~ Bldg-Alter~
##  8 1318 E 7T~ Issued        Rough           Partial Inspecti~ CofO ~ Bldg-Alter~
##  9 1318 E 7T~ Issued        Rough           Partial Inspecti~ Issued Bldg-New
## 10 1318 E 7T~ Issued        Rough           Partial Inspecti~ Issued Bldg-Alter~
## 11 3911 S FI~ Issued        Footing/Founda~ Approved          Permi~ Electrical
## 12 225 E 31S~ Issued        Green Building~ Not Ready for In~ Issued Bldg-Addit~
## 13 1026 S BR~ Issued        Rough           Insp Scheduled    Permi~ Bldg-Demol~
## 14 1026 S BR~ Issued        Rough           Insp Scheduled    Permi~ Bldg-Alter~
## 15 1026 S BR~ Issued        Rough           Insp Scheduled    Permi~ Bldg-Demol~
## # i 6 more variables: contractor_city <chr>, contractor_state <chr>,
## #   applicant_address_3 <chr>, zone <chr>, contractor_city_binary <dbl>,
## #   inspection_binary <dbl>
```

apply chi square test to see if there is a relationship between inspection result and contractor city

```
# Create a contingency table
contingency_table = table(bldg_safty_insp_permit_2_LA_3_insp_binary$contractor_city,
                          bldg_safty_insp_permit_2_LA_3_insp_binary$inspection_binary)
print(contingency_table) %>% head()
```

```
##
##                                 0     1
##   \\LAKE FOREST              11     6
##   \\PASADENA                 32    17
##   5BELL CANYON                6     6
##   ACTON                     196    86
##   AGOURA                     50    22
##   AGOURA HILLS              294   114
##   AGUA DULCE                199    64
##   ALAMED                      1     1
##   ALAMEDA                   938   235
##   ALBUQUERQUE                 2     2
##   ALHAMBRA                    0    10
##   ALISO VIEJO                19    10
##   ALPHARETTA                  5     2
##   ALPINE                      9    11
##   ALTA LOMA                 141    51
##   ALTADENA                  376   136
##   ANAHEIM                  6177  1766
##   ANAHEIM HILLS             589   152
##   ANTIOCH                     1     1
##   APPLE VALLEY                4     1
##   APTOS                       4     2
##   ARCADIA                   374   133
##   ARLETA                     85    49
##   ARLINGTON                   2     1
##   ARROYO GRANDE             134    28
##   ARTESIA                    139    54
##   ARVADA                     254    84
##   ATASCADERO                   9     2
##   ATLANTA                    367   105
##   AVILA BEACH                 50    22
##   AZUSA                      533   173
##   BAKERSFIELD                248    78
##   BALDWIN PARK               112    54
##   BANNING                      2     1
##   BATON ROUGE                107    14
##   BELL                         8     7
##   BELL CANYON                 57    27
##   BELL GARDENS                32    12
##   BELLFLOWER                  49    33
##   BETHESDA                  1091   292
##   BEVERLY HILLS             2667   995
##   BLUE JAY                    20     6
##   BONSALL                     25    11
##   BOSTON                     357   155
##   BREA                      1018   369
##   BRENTWOOD                    4     2
##   BROOKINGS                    6     2
##   BUENA PARK                 483   108
##   BURBANK                   1855   478
##   CALABASAS                 2046   661
##   CALABASAS HILLS             20    16
##   CALIMESA                    39    11
##   CAMARILLO                  244   106
##   CANOGA PARK                520   172
##   CANYON COUNTRY             257    65
##   CARLSBAD                   397   152
##   CARSON                    1312   532
##   CASTAIC                    100    53
##   CATHEYS VALLEY               4     6
##   CERRITOS                   122    88
##   CHATSWORTH                1478   508
##   CHINO                      370   110
##   CHINO HILLS                470   172
##   CHULA VISTA                 78    51
##   CITRUS HEIGHTS              27     6
##   CITY OF COMMERCE            39    12
##   CITY OF INDUSTRY           673   223
##   CLAREMONT                  435    94
##   CLEARWATER                  32    17
##   COLTON                      30     8
##   COMMERCE                   156    78
##   COMPTON                     33    12
##   CONCORD                    844   281
##   COQUITLAM B C               95    35
##   CORONA                     899   311
##   CORONA, CA                  95    35
##   COSTA MESA                 515   165
##   COVINA                     773   253
##   CROWLEY                    103    26
##   CULVER CITY               635   243
##   CYPRESS                   315    91
##   DALLAS                    210    63
##   DEL MAR                     4     2
##   DENVER                     12    14
##   DEPERE                      8    12
##   DIAMOND BAR               207    43
##   DOWNEY                    297   183
##   DUARTE                    192    36
##   EAST SYRACUSE              66    23
##   EASTVALE                    2     1
##   EL CAJON                   57    18
##   EL MONTE                  103    26
##   EL SEGUNDO                230    45
##   ELIZABETHTOWN             188    64
##   ENCINITAS                 253    99
##   ENCINO                   1356   491
##   ESCONDIDO                 184    73
##   FAIRFIELD                 209    64
##   FLUSHING                  102    32
##   FONTANA                   452   160
##   FORT LAUDERDALE             0     1
##   FOUNTAIN VALLEY           709   213
##   FULLERTON                2207   695
##   GARDEN GROVE             1004   323
##   GARDENA                   813   312
##   GLENDALE                 3556  1187
##   GLENDORA                  182    59
##   GOLDSBORO                  96    32
##   GRANADA HILLS             183    77
##   GREELEY                   480   228
##   GUASTI                    144    35
##   HACIENDA HEIGHTS           13    17
##   HARBOR CITY                20    16
##   HAWAIIAN GARDENS           26    12
##   HAWTHORNE                 227    85
##   HAYWARD                   103    26
##   HEMET                     191    71
##   HENDERSON                   2     0
##   HERMOSA BEACH             122    33
##   HESPERIA                   30     5
##   HIGHLAND                   32    12
##   HOLLYWOOD                 854   297
##   HOUSTON                   691   210
##   HUNTINGTON BEACH         1323   430
##   HUNTINGTON PARK            12     9
##   IMPERIAL                   13     6
##   INGLEWOOD                 431   148
##   IRVINE                   4629  1448
##   IRWINDALE                 422   150
##   JURUPA VALLEY              35     3
##   KALISPELL                   1     0
##   KEENE                      16     7
##   KENNESAW                    1     0
##   LA CANADA                 122    44
##   LA CANADA FLINTRIDGE       12     9
##   LA CRESCENTA              336   138
##   LA HABRA                   36    16
##   LA HABRA HEIGHTS           27    19
##   LA MIRADA                   9     3
##   LA PALMA                   13     6
##   LA PUENTE                  95    63
##   LA VERNE                  370   125
##   LADERA RANCH                2     1
##   LAGUNA BEACH              695   218
##   LAGUNA HILLS              135    63
##   LAGUNA NIGUEL              74    28
##   LAKE ELSINORE               4     6
##   LAKE FOREST               323   117
##   LAKE HUGHES                 3     1
##   LAKE VIEW TERRACE          17    17
##   LAKESIDE                   95    35
##   LAKEWOOD                    5     2
##   LANCASTER                 170    88
##   LAS VEGAS                  97    39
##   LAWNDALE                   29    16
##   LIBERTYVILLE               27     4
##   LITCHFIELD PARK             3     1
##   LITTLE ROCK               149    46
##   LIVERMORE                 189    67
##   LOGANVILLE                  0     1
##   LOMITA                    944   782
##   LONG BEACH               2244   752
##   LOS ALAMITOS              416   166
##   LOS ANGELES             60842 28711
##   LOS ANGELES,               12     4
##   LOS GATOS                 171    55
##   LYNWOOD                    66    35
##   MALIBU                     14    18
##   MANHATTAN BEACH            34    11
##   MARINA DEL REY            223    68
##   MARTINEZ                 3942   675
##   MENLO PARK                 84    33
##   MERCER ISLAND             107    28
##   MIDDLETOWN                 44    13
##   MINNEAPOLIS                95    35
##   MINNETONKA                 25    11
##   MIRA LOMA                 218    67
##   MISSION HILLS               4     2
##   MISSION VIEJO             416   145
##   MONROVIA                  907   281
##   MONSEY                    105    32
##   MONTCLAIR                   0     1
##   MONTEBELLO                 44    30
##   MONTEREY                   18    27
##   MONTEREY PARK            1093   405
##   MONTROSE                  825   265
##   MONTROSE, CA 91020         19     4
##   MONUMENT                   70     6
##   MOORPARK                   57    32
##   MORENO VALLEY               0     2
##   MORRISTOWN               3332   896
##   MURRIETA                   52    19
##   N HOLLYWOOD               126    42
##   NEW YORK                  797   232
##   NEWBURY PARK              404   114
##   NEWHALL                   253    71
##   NEWPORT BEACH            1498   389
##   NEWPORT COAST             191    70
##   NORCO                      87    34
##   NORTH HILLS                91    63
##   NORTH HOLLYWOOD           990   323
##   NORTHRIDGE                374   167
##   NORWALK                  1475   783
##   NOVATO                     50    22
##   OAKDALE                     2     3
##   OAKLAND                    21    59
##   OCEANSIDE                 156    69
##   OJAI                       26     8
##   ONTARIO                   524   149
##   ORANGE                   3253   809
##   OXNARD                    347   105
##   PACIFIC PALISADES          29    20
##   PACOIMA                    32     3
##   PALM SPRINGS                0     4
##   PALMDALE                   15    16
##   PANORAMA CITY              64    48
##   PARAMOUNT                 613   195
##   PARSIPPANY                515   158
##   PASADENA                 2755  1029
##   PATTON                      0     1
##   PATTON, CA                  0     1
##   PEARBLOSSOM                26    16
##   PERRIS                     92    29
##   PHEONIX                   103    26
##   PHOENIX                   205    82
##   PICO RIVERA                20    31
##   PINE MOUNTAIN CLUB        276    96
##   PISMO BEACH               268    56
##   PLACENTIA                 665   226
##   PLAYA DEL REY               7     2
##   POMONA                    226    61
##   PORTER RANCH                2     1
##   PORTLAND                  103    26
##   POWAY                       9     5
##   PRIOR LAKE                 19     4
##   QUARTZ HILL                21     5
##   QUINCY                     14     1
##   RAMONA                     14     1
##   RANCHO CUCAMONGA          348   125
##   RANCHO DOMINGUEZ         2756   838
##   RANCHO PALOS VERDES       219    95
##   RANCHO SANTA MARGARITA    252    76
##   REDLANDS                  204    71
##   REDONDO BEACH             792   263
##   REDWOOD CITY               11     4
##   REEDLEY                     6     3
##   RESEDA                    311   119
##   RIALTO                     25    18
##   RICHARDSON                286    56
##   RICHMOND                   75    23
##   RIVERSIDE                 729   212
##   ROSEMEAD                  259   109
##   ROWLAND HEIGHTS            42    26
##   SACHSE                    103    26
##   SACRAMENTO                555   139
##   SALT LAKE CITY             80    19
##   SAN BERNARDINO            162    70
##   SAN CARLOS                264    66
##   SAN CLEMENTE              250   108
##   SAN DIEGO                1415   393
##   SAN DIMAS                1797   580
##   SAN FERNANDO             1133   316
##   SAN FRANCISCO            5359  1735
##   SAN FRANCISCO,             90    23
##   SAN FRANCISCO, CA 94107   103    26
##   SAN FRANSICO                9     5
##   SAN GABRIEL                10     3
##   SAN JOSE                 1259   325
##   SAN JUAN BAUTISTA          66    11
##   SAN JUAN CAPISTRANO        96    35
##   SAN MARCOS                373   107
##   SAN PEDRO                 331    83
##   SANTA ANA                1078   367
##   SANTA CLARITA            1143   368
##   SANTA FE SPRINGS         3151   971
##   SANTA MONICA             2642   804
##   SARATOGA                   26    22
##   SAUGUS                     58    14
##   SCOTTSDALE                545   155
##   SEAL BEACH                 32    17
##   SEATTLE                   279    84
##   SEYMOUR                     4     1
##   SHADOW HILLS               12     4
##   SHERMAN OAKS             1660   565
##   SIERRA MADRE               25    11
##   SIGNAL HILL               546   183
##   SIMI VALLEY              1664   600
##   SO EL MONTE               235    56
##   SOMIS                       1     1
##   SOUTH EL MONTE            174    59
##   SOUTH GATE                502   164
##   SOUTH PASADENA            184    56
##   ST JOSEPH                 108    39
##   ST LOUIS                  103    26
##   ST PETERSBURG             115    24
##   STANTON                    28     2
##   STEVENSON RANCH            85    28
##   STREETSBORO                52    32
##   STUDIO CITY               561   336
##   STURTEVANT                218    67
##   SUN VALLEY               1241   375
##   SUNLAND                    42    18
##   SUNNYVALE                  32    17
##   SYLMAR                   1671   466
##   TARZANA                   888   313
##   TEMECULA                  303    48
##   TEMPLE CITY                22    17
##   THOUSAND OAKS            1398   461
##   THOUSAND PALMS              2     1
##   TOLUCA LAKE                78    82
##   TORRANCE                 1763   547
##   TRACY                      94    32
##   TUJUNGA                   529   242
##   TURLOCK                    14     1
##   TUSTIN                    626   174
##   UPLAND                    201    83
##   VALENCIA                 1182   422
##   VALLEY CENTER               2    10
##   VALLEY VILLAGE             79    63
##   VAN NUYS                 1992   690
##   VENICE                    118    37
##   VENTURA                   529   127
##   VERDUGO CITY               10    25
##   VERNON                     32    30
##   VICTORVILLE                95    37
##   VISTA                       2     1
##   WALNUT                    262    90
##   WALTHAN                    26    17
##   WATSONVILLE                33    15
##   WEST COVINA               446   123
##   WEST HILLS                755   240
##   WEST HOLLYWOOD             90    54
##   WEST LOS ANGELES           60    11
##   WEST SACRAMENTO           288    54
##   WESTCHESTER                 0     1
##   WESTLAKE VILLAGE          651   231
##   WESTMINSTER               155    60
##   WHITTIER                  247    78
##   WILMINGTON                362    99
##   WINNETKA                  121    59
##   WINTER GARDEN              95    35
##   WOODLAND HILLS           3623  1301
##   YORBA LINDA               853   149
```

```
##
##                      0   1
##   \\LAKE FOREST     11   6
##   \\PASADENA        32  17
##   5BELL CANYON       6   6
##   ACTON         196  86
##   AGOURA         50  22
##   AGOURA HILLS  294 114
```

```
chi_test_result = chisq.test(contingency_table)
```

```
## Warning in chisq.test(contingency_table): Chi-squared approximation may be
## incorrect
```

```
print(chi_test_result)
```

```
## 
##  Pearson's Chi-squared test
## 
## data:  contingency_table
## X-squared = 4084.3, df = 342, p-value < 2.2e-16
```

It shows that the contractor from Los Angeles has a higher chance of getting a successful inspection result.

But it hightly depends on how the variables were converted to binary. It will need to be checked with the domain expert.