Contents

List of Tables

List of Figures

```r
# Ensure plyr is detached to avoid conflicts
if ("package:plyr" %in% search()) {
  detach("package:plyr", unload = TRUE)
}

# Ensure datetime is in proper format
weather_data <- weather_data %>%
  mutate(datetime = as.POSIXct(datetime, format = "%Y-%m-%d %H:%M:%S"))

# Combine rows by month
weather_data_processed <- weather_data %>%
  mutate(month = format(as.Date(datetime), "%Y-%m")) %>% # Extract month as "YYYY-MM"
  group_by(month) %>%
  summarise(
    avg_temperature = mean(hourly_temperature_2m, na.rm = TRUE),
    max_temperature = max(hourly_temperature_2m, na.rm = TRUE),
    min_temperature = min(hourly_temperature_2m, na.rm = TRUE),
    total_records = dplyr::n() # Explicitly use n() from dplyr
  )

weather_data_processed
```

```
## # A tibble: 25 x 5
##    month   avg_temperature max_temperature min_temperature total_records
##    <chr>             <dbl>           <dbl>           <dbl>         <int>
##  1 2020-01            7.59            21.2            -4.9           739
##  2 2020-02            8.44            20.2            -6.3           696
##  3 2020-03           13.3             29.4            -4.3           745
##  4 2020-04           15.1             29.6             2.9           720
##  5 2020-05           18.6             32.4             3.5           744
##  6 2020-06           23.7             34.2            13.8           720
##  7 2020-07           28.1             37.4            20.5           744
##  8 2020-08           25.5             34.5            17.8           744
##  9 2020-09           21.1             34.5             7.2           720
## 10 2020-10           17.2             27.8             6.1           744
## # i 15 more rows
```

Rationale and Research Questions

```r
# Process the data: Combine rows by month
Stanislaus_Air_Data_Processed <- Stanislaus_Air_Data %>%
  mutate(
    Date = as.Date(Date, format = "%m/%d/%Y"),
    month = format(Date, "%Y-%m")
  ) %>%
  group_by(month) %>%
  summarise(
    avg_PM2_5 = mean(Daily.Mean.PM2.5.Concentration, na.rm = TRUE),
    max_PM2_5 = max(Daily.Mean.PM2.5.Concentration, na.rm = TRUE),
    min_PM2_5 = min(Daily.Mean.PM2.5.Concentration, na.rm = TRUE),
    total_records = n()
  )


Stanislaus_Air_Data_Processed
```

```
## # A tibble: 132 x 5
##    month   avg_PM2_5 max_PM2_5 min_PM2_5 total_records
##    <chr>       <dbl>     <dbl>     <dbl>         <int>
##  1 2013-01      27.0      57.7       1.8            67
##  2 2013-02      16.9      41.3       3.1            60
##  3 2013-03       8.94     24.7       1.9            67
##  4 2013-04       5.74     12.6      -0.5            64
##  5 2013-05       8.50     24.5       2.1            70
##  6 2013-06       6.44     12.1       2              65
##  7 2013-07      10.6      33.4       5              69
##  8 2013-08       7.22     20.4       0.2            65
##  9 2013-09       5.73     10.9      -1.2            63
## 10 2013-10      12.3      21.9       4.2            65
## # i 122 more rows
```

Dataset Information

We are download two dataset, one from EPA Stanislaur_Air_Data to present the air quality(include two values we want to examine PM2.5 and AQI values), another is weather data from NOAA, we installed openmeteo" package to process data. Since two datasets have different format of dates, we first use mutate function and group _by month to combine the rows for ensuring the consistency of "yyyy%-mm%" format. However, each month has differnt numbers of observations, we then combine all the rows happen in the same month so that we can easier to analyze the data.

weather_data to see how

Exploratory Analysis

Analysis

Question 1: <insert specific question here and add additional subsections for additional questions below, if needed>

Question 2:

Summary and Conclusions

References