**RAND CORPORATION**

**Air Quality Analysis One**

Johnathan Lovett

12 August 2019

**Table of Contents:**

1. Introduction………………………………………………………………………………………..3
2. Analysis of Zip Code Random Week 10/08/18 through 10/14/18……...…………………………4
   1. Zip Code Data Visualization.……………………………………………………………..5
3. Analysis of PM (1.0, 2.5, 10.0) by hour of random week: 10/08/18 through 10/14/18………….13
   1. Hourly Data Visualization……………….…………………..…………………………..14
4. Descriptive Statistics………..………………………………………….…………………………21
5. Moving Forward: Analyzing July 2018 through July 2019…………..….……………………….23
6. For Reference: Temperature and Humidity Graphs of Week 10/08/18 through 10/14/18……….26

**1. Introduction:**

* This analysis isolates a single, randomized week to experiment with data analysis and visualization as a start - that may be extracted to analyze the entire data set.
* The week 10/08/2018 through 10/14/2018 in Santa Monica was randomly selected, and the following metrics are examined: how area, zip code, day of the week/weekend, and time of day factor into air quality, as measured through changes in Particulate Matter 1.0, 2.5, 10.0 (ug/m3), Temperature, and Humidity.
* I intend to apply the randomized October week analysis to entire July 2018 through July 2019 data set, month by month. I intend to begin a Machine Learning Python & R certificate [course](https://www.udemy.com/machinelearning/?ranMID=39197&ranEAID=vedj0cWlu2Y&ranSiteID=vedj0cWlu2Y-7iYKKrsTpEhpNAttwFgbcg&LSNPUBID=vedj0cWlu2Y) this week to supplement analytical tools.

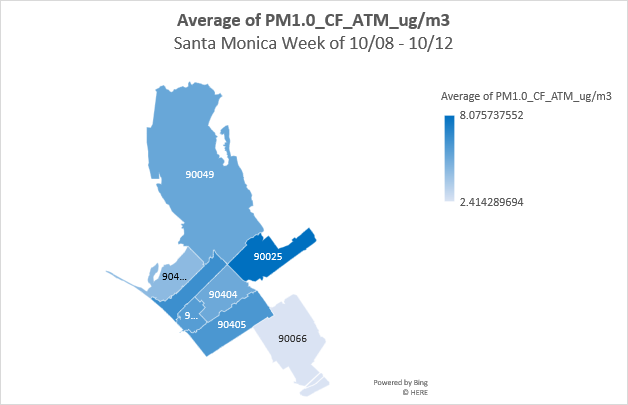
**2. Sample Analysis of a Random week: 10/08 through 10/14, October 2018:**

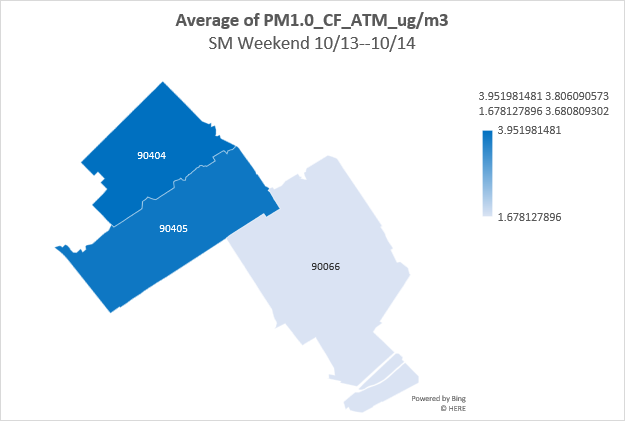
* During week of tested locations, 90025 Zip Code is most affected by 1.0, 2.5, and 10.0 Particulate Matter sizes and is of highest temperature (area not tested on weekend).
* Friday, Tuesday, Sunday (descending order) days are most affected by 1.0 PM and 2.5 PM. Friday had the second lowest humidity and temperature of the week.
* Tuesday, Friday, Sunday (descending order) days are most affected by 10.0PM.
* Saturday, Thursday (ascending order) witness the lowest levels of all 1.0, 2.5, 10.0 PMs.
* Monday and Wednesday are of nearly equivalent levels for 1.0, 2.5, 10.0 PMs.
* Of tested zip codes on weekend (90404, 90405, 90066), 90404 is most affected by 1.0, 2.5, 10.0 PMs.

90404 witnesses the lowest temperature yet is most humid.

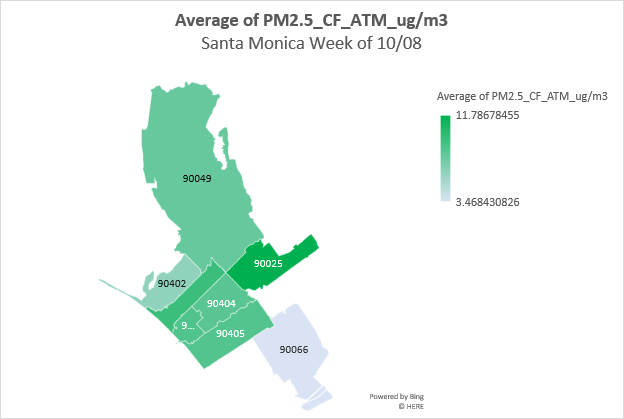
* + The uptick in PM on Friday and Sunday is explained by the increased volume of commuting cars to and from Santa Monica (i.e. tourism to 90404 beach-adjacent zip code) and far less tourism to the more inland Mar Vista (90066).
  + The Tuesday uptick is possibly unique to this week, though further weeks ought to be identified if there’s a trend.

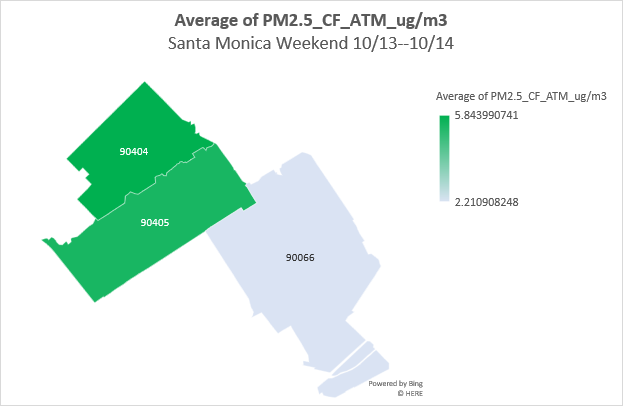
**10/08 – 10/14 Average Weekday vs Average Weekend Particulate Matter 1.0:**



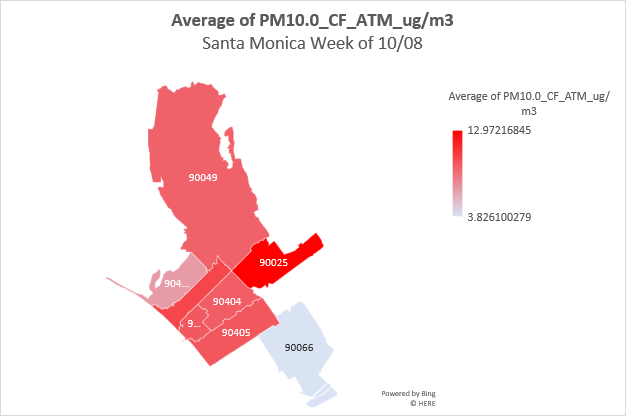


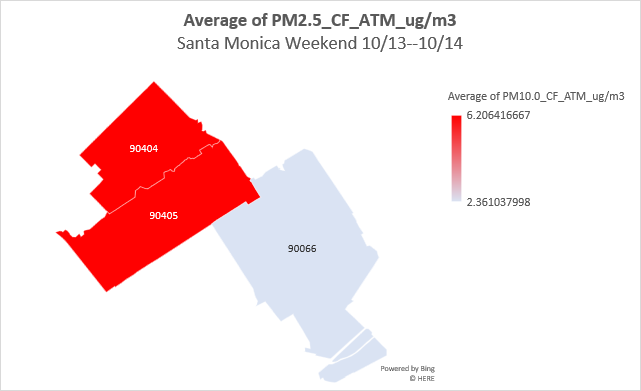
**10/08 – 10/14 Average Weekday vs Average Weekend Particulate Matter 2.5:**



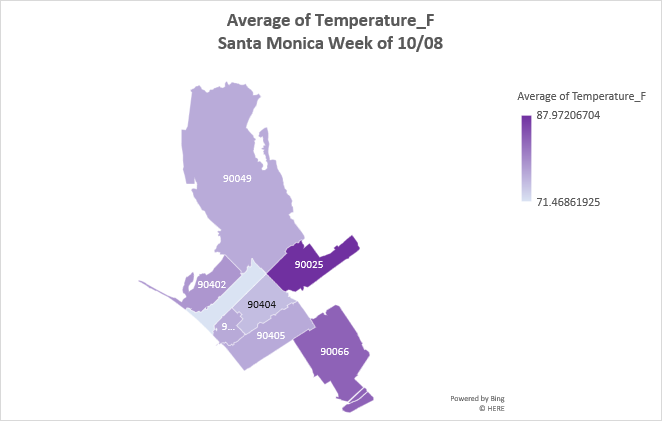


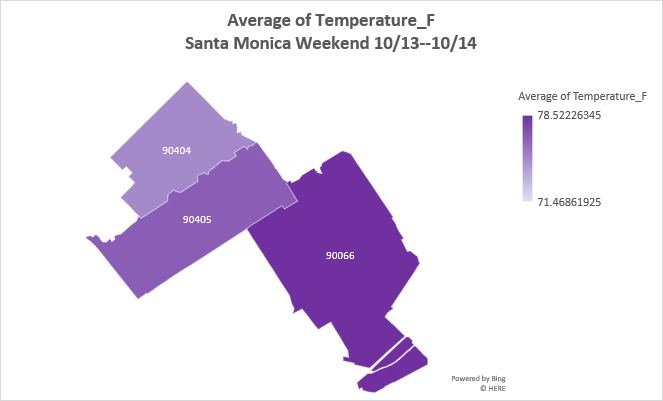
**10/08 – 10/14 Average Weekday vs Average Weekend Particulate Matter 10.0:**



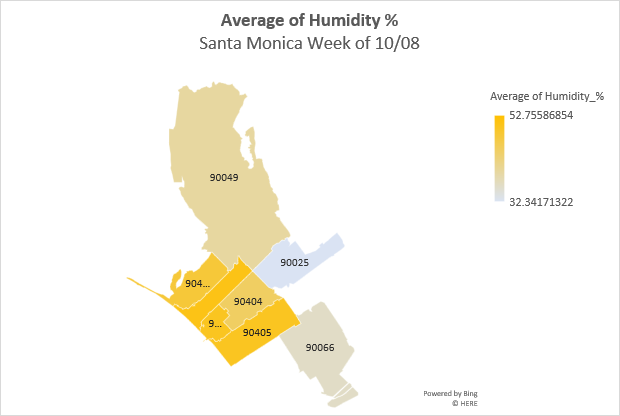


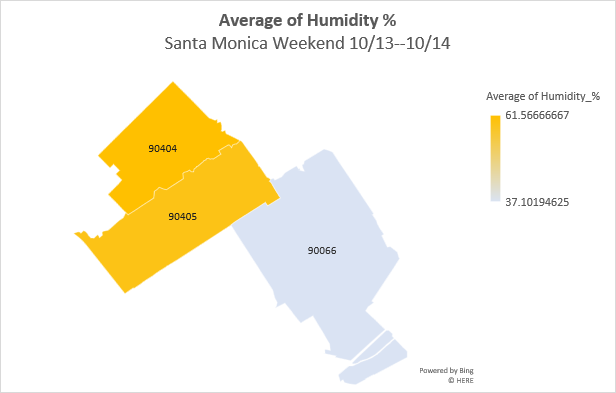
**10/08 – 10/14 Average Weekday vs Average Weekend Temperature:**





**10/08 – 10/14 Average Weekday vs Average Weekend Humidity %:**





**3. Sample Analysis of PM (1.0, 2.5, 10.0) by hour of random week: 10/08/18 through 10/14/18:**

* For this week, generally behavior of one size of particulate matter is predictive of the other two.
* On Monday for Zip Code 90401 (ThingSpeakPrimaryId 395536), time ranges of highest activity of 1.0, 2.5, 10.0 PM’s occurred between 00:00-02:20 and 16:25-23:24 (ascending order).
* On Tuesday for Zip Code 90402 (ThingSpeakPrimaryId 569301), time range 08:36-20:16 witnessed highest levels of 1.0, 2.5, 10.0 PM’s, peaking around 17:09. This is roughly the case for Zip Codes 90405, 90404, and 90025 (90025 features an amplitude that is about half the others).
* On Wednesday for Zip Code 90066, 1.0, 2.5, 10.0 PM’s all dramatically increased from effectively 0 to over 200 CF ATM ug/m3 at around 21:53, decreasing until about 23:47.
* On Thursday, similar sudden spikes in 1.0, 2.5, 10.0 PM’s occurred in zip codes 90404 at 02:20 and 90049 at around 04:00.
* On Friday, Zip Code 90066 peaked at 21:58, 04:34, and 03:03 in descending order in in 1.0, 2.5, 10.0 PM’s.
* On Saturday Zip Code 90404 peaked at around 17:02 until 18:36, and less intensely at 23:16.
* Sunday Zip Code 90066 and 90403 peaked 06:19 and 19:32, and 90404 peaked suddenly at 00:11. These

**10/08-10/14 PM (1.0, 2.5, 3.0) by Hour**

**4. Descriptive Statistics**

|  |  |  |  |  |
| --- | --- | --- | --- | --- |
| **Santa Monica Air Quality Data Analysis** | | | | |
|  |  |  |  |  |
| **PM1.0\_CF\_ATM\_ug/m3** | | | **PM2.5\_CF\_ATM\_ug/m3** | |
|  |  |  |  |  |
| Mean | 8.268892 |  | Mean | 13.06398 |
| Standard Error | 0.014125 |  | Standard Error | 0.023575 |
| Median | 4.77 |  | Median | 7.11 |
| Mode | 1 |  | Mode | 0.11 |
| Standard Deviation | 14.46236 |  | Standard Deviation | 24.13916 |
| Sample Variance | 209.1597 |  | Sample Variance | 582.6988 |
| Kurtosis | 217.7524 |  | Kurtosis | 625.0895 |
| Skewness | 10.71164 |  | Skewness | 14.94293 |
| Range | 617.54 |  | Range | 2497.25 |
| Minimum | 0 |  | Minimum | 0 |
| Maximum | 617.54 |  | Maximum | 2497.25 |
| Sum | 8669032 |  | Sum | 13696159 |
| Count | 1048391 |  | Count | 1048391 |
| Confidence Level(95.0%) | 0.027684 |  | Confidence Level(95.0%) | 0.046207 |
|  |  |  |  |  |
| **PM10.0\_CF\_ATM\_ug/m3** | | | **UptimeMinutes** | |
|  |  |  |  |  |
| Mean | 14.69179 |  | Mean | 2793.829 |
| Standard Error | 0.027551 |  | Standard Error | 5.688051 |
| Median | 7.95 |  | Median | 801 |
| Mode | 1 |  | Mode | 0 |
| Standard Deviation | 28.20938 |  | Standard Deviation | 5824.562 |
| Sample Variance | 795.7692 |  | Sample Variance | 33925522 |
| Kurtosis | 853.8595 |  | Kurtosis | 18.45844 |
| Skewness | 16.81847 |  | Skewness | 3.948633 |
| Range | 3383.65 |  | Range | 54583 |
| Minimum | 0 |  | Minimum | 0 |
| Maximum | 3383.65 |  | Maximum | 54583 |
| Sum | 15402740 |  | Sum | 2.93E+09 |
| Count | 1048391 |  | Count | 1048575 |
| Confidence Level(95.0%) | 0.053998 |  | Confidence Level(95.0%) | 11.14839 |
|  |  |  |  |  |
|  |  |  |  |  |
| **ADC** |  |  | **Temperature\_F** | |
|  |  |  |  |  |
| Mean | -62.0922 |  | Mean | 75.2907 |
| Standard Error | 0.020944 |  |  |  |
| Median | -64 |  | \*Some flawed data | |
| Mode | 31 |  | showed temp as | |
| Standard Deviation | 21.44681 |  | exponentially large | |
| Sample Variance | 459.9657 |  | so generated mean | |
| Kurtosis | 8.994232 |  | manually. | |
| Skewness | 2.562871 |  |  |  |
| Range | 128 |  |  |  |
| Minimum | -97 |  |  |  |
| Maximum | 31 |  |  |  |
| Sum | -6.5E+07 |  |  |  |
| Count | 1048575 |  |  |  |
| Confidence Level(95.0%) | 0.04105 |  |  |  |
|  |  |  |  |  |
| **Humidity\_%** | |  | **PM2.5\_CF\_1\_ug/m3** | |
|  |  |  |  |  |
| Mean | 44.95242 |  | Mean | 11.83538 |
| Standard Error | 0.015858 |  | Standard Error | 0.016831 |
| Median | 45 |  | Median | 7.11 |
| Mode | 42 |  | Mode | 0.11 |
| Standard Deviation | 16.12178 |  | Standard Deviation | 17.2351 |
| Sample Variance | 259.9118 |  | Sample Variance | 297.0487 |
| Kurtosis | 57.44505 |  | Kurtosis | 463.9136 |
| Skewness | 4.334535 |  | Skewness | 11.86733 |
| Range | 255 |  | Range | 1663.8 |
| Minimum | 0 |  | Minimum | 0 |
| Maximum | 255 |  | Maximum | 1663.8 |
| Sum | 46459046 |  | Sum | 12410038 |
| Count | 1033516 |  | Count | 1048554 |
| Confidence Level(95.0%) | 0.031082 |  | Confidence Level(95.0%) | 0.0329888A1:E71390430238 |

**5. Moving Forward: Beginning to Analyze by Month, July 2018 through July 2019**

\*Including sample data I’ve begun to work with.

In July 18, zip Code 90066 had highest levels of 1.0 PM, peaking the 28th at nearly 250 CF\_ATM\_ug/m3.

\*Not including every month or chart to not overwhelm length.

**6. For reference: Analysis of 10/08-10/14 Temperature and Humidity by Hour:**