Middle English: from Latin *luna* - ‘moon’

From the belief that changes of the moon caused intermittent insanity.

LUNATIC

DOES THE DATA BACKUP THE OLD SAYINGS

CAN WE USE THE WEATHER FORECAST TO PREDICT UPCOMING CRIME LEVELS

CRIME AND WEATHER IN SAN FRANCISCO

Jim Haskin - GA Data Science - December 2016

HISTORICAL CRIME INFORMATION

Incident reports of the San Francisco Police Department from the SF OpenData website. From 2003 until early 2016. 1,866,570 incidents.

Created Daily Crime Level Statistic

	Category	Descript	DayOfWeek	Date	Time	PdDistrict	Resolution	Address	X	Y	Location	PdId
IncidntNum												
160051264	WARRANTS	WARRANT ARREST	Monday	01/18/2016	23:52	CENTRAL	ARREST, BOOKED	400 Block of POWELL ST	-122.408568	37.788759	(37.7887594214703, -122.408568445228)	16005126463010
160051242	ROBBERY	ROBBERY, BODILY FORCE	Monday	01/18/2016	23:40	TENDERLOIN	NONE	100 Block of STOCKTON ST	-122.406428	37.787109	(37.78710945429, -122.40642786236)	16005124203074

PREDICTION VARIABLES – POSSIBILITIES

- All Crime
- Crime I believe are effected by weather
- Violent crimes
- Location
- Time of Day

PREDICTION VARIABLES – CRIME CATEGORIES

larceny/theft	370337
other offenses	260051
non-criminal	193041
assault	160146
vehicle theft	110911
drug/narcotic	110273
vandalism	92436
warrants	88371
burglary	76509
suspicious occ	64123
missing person	53887
robbery	47879
fraud	31146
secondary codes	20767
forgery/counterfeiting	18487
weapon laws	18159
trespass	15445
prostitution	15347
stolen property	9815
drunkenness	8894
disorderly conduct	8799
sex offenses, forcible	8471
recovered vehicle	6318

PREDICTION VARIABLES – DESCRIPTION – ASSAULT

battery	57142
threats against life	29896
inflict injury on cohabitee	14842
aggravated assault with a deadly weapon	13584
aggravated assault with bodily force	10382
battery, former spouse or dating relationship	5990
aggravated assault with a knife	5134
battery of a police officer	2802
child abuse (physical)	2652
aggravated assault with a gun	2127
threatening phone call(s)	1729
battery with serious injuries	1697
stalking	1619
elder adult or dependent abuse (not embezzlement or theft)	1286
assault	1092
assault with caustic chemicals	887
false imprisonment	805
attempted simple assault	710
attempted homicide with a gun	618
shooting into inhabited dwelling or occupied vehicle	540
assault on a police officer with a deadly weapon	489

PREDICTION VARIABLES

Crime Count

Count of all Crimes

Violent Count

Count of Crimes from the assault, rape and secondary categories

Crimes Of Passion (COP) Count

```
cop_words = ['assault', 'battery', 'drunk', 'abuse', 'forced', 'rape', 'shooting',
             'violence', 'harassing', 'threat', 'threatening', 'threats', 'resist', 'resisting',
             'destruction', 'weapons', 'gun', 'knife', 'armed', 'deadly', 'drunkenness',
             'bomb', 'bombing', 'influence', 'looting', 'disorderly', 'force', 'forcible',
             'fighting', 'injuries', 'nuisance', 'homicide', 'alcohol', 'rape', 'mayhem',
             'abuse', 'cruelty', 'lewd', 'molest', 'disturbing']
```

HISTORICAL WEATHER INFORMATION

Historical data from Weather Underground from 2003 until 2015.

FieldName	Type	Description
pst	string	Date in format : 2003-1-1
max_temperaturef	int	High Temperature for the day in degrees F
mean_temperaturef	int	Mean Temperature for the day in degrees F
min_temperaturef	int	Low Temperature for the day in degrees F
max_dew_pointf	int	High Dew Point for the day in degrees F
meandew_pointf	int	Mean Dew Point for the day in degrees F
min_dewpointf	int	Low Dew Point for the day in degrees F
max_humidity	int	Maximum Humidity for the day in percentage
mean_humidity	int	Mean Humidity for the day in percentage
min_humidity	int	Minimum Humidity for the day in percentage
max_sea_level_pressurein	float	High Sea Level for the day in inches
mean_sea_level_pressurein	float	Mean Sea Level for the day in inches
min_sea_level_pressurein	float	Low Sea Level for the day in inches
max_visibilitymiles	int	Maximum Visibility in miles
mean_visibilitymiles	int	Mean Visibility in miles
min_visibilitymiles	int	Minimum Visibility in miles
max_wind_speedmph	int	Maximum maintained Wind Speed in mph

“THE COLDEST WINTER I EVER SPENT WAS A SUMMER IN SAN FRANCISCO.”

–Mark Twain ?

FEATURE ENGINEERING

- pywws.conversions Modual

  Wind Chill - Temperature and Wind

  Apparent Temperature - Temperature, Humidity and wind

  Heat Index - Temperature, Humidity and Dew Point

  Dew Point - Temperature and Humidity

- ephem module in a script (sf_sun_moon.py)

  Sun Length - Daily Minutes the Sun is up

  Sun Level - Cloud Cover and Length of Sunshine

  Moon Phase

WEATHER FORECAST

- Weather Underground - json - request

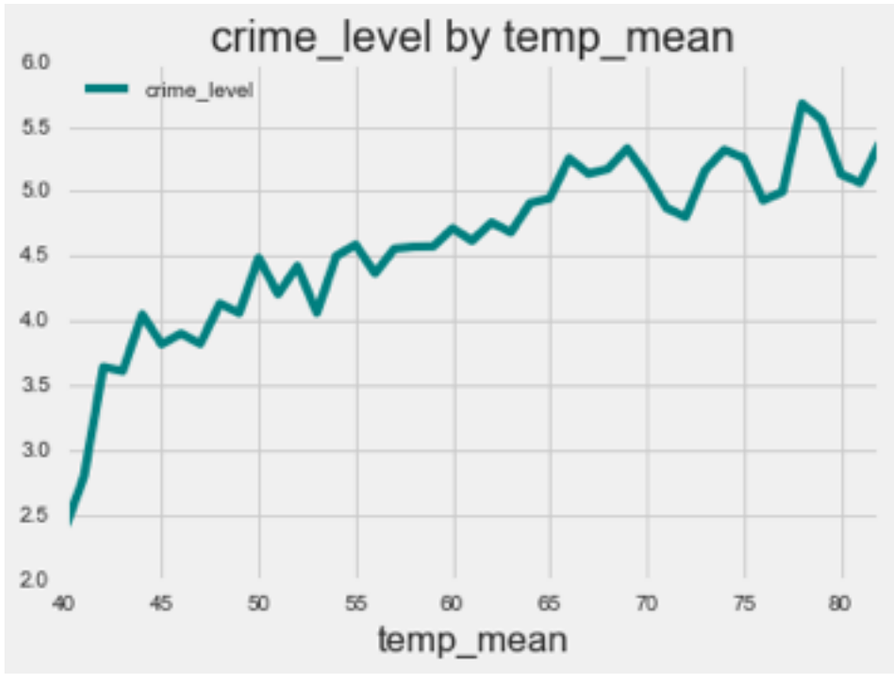
```
'date', 'dayofweek', 'day', 'month', 'year',
'temp_max', 'temp_mean', 'temp_min', 'temp_delta',
'humidity_mean',
'wind_speed_max', 'wind_speed_mean',
'precipitation', 'events', 'wind_direction']
```

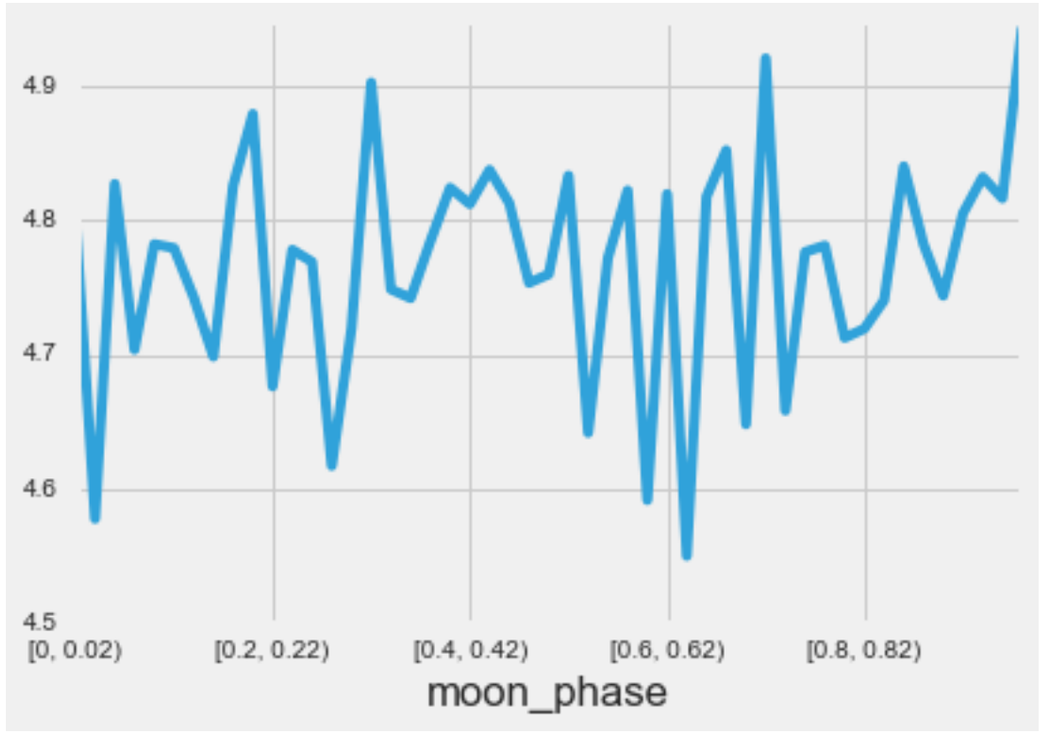
- Open Weather Map - pyowm wrapper

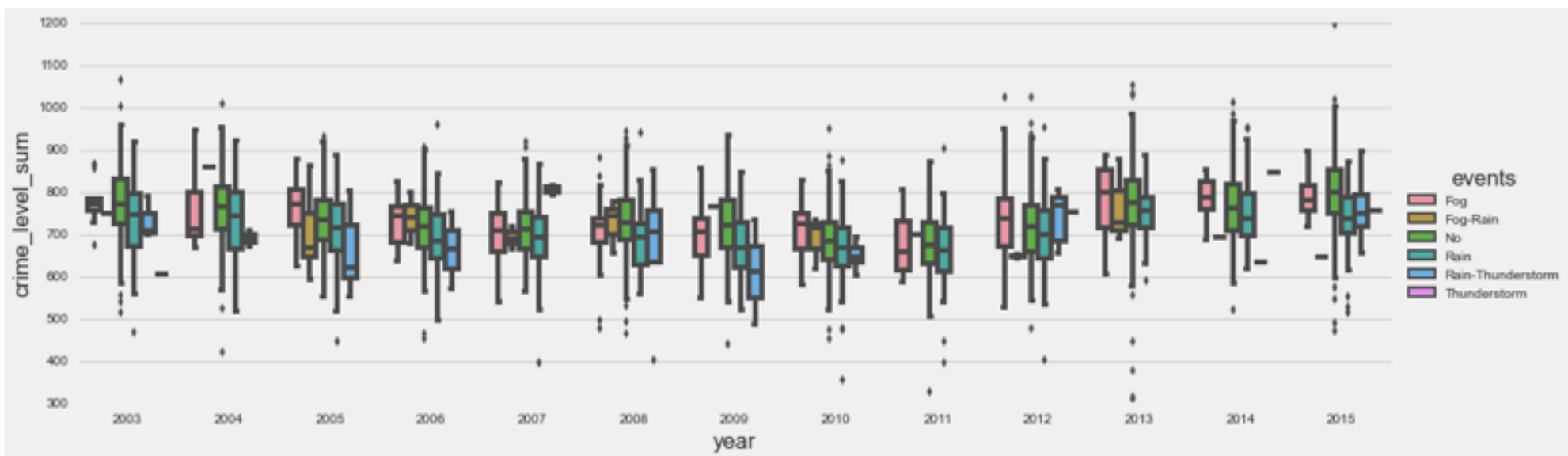
```
['date', 'sl_pressure_mean', 'cloud_cover', 'owm_status']
```

- Engineered Features added - Same as Modeling Data

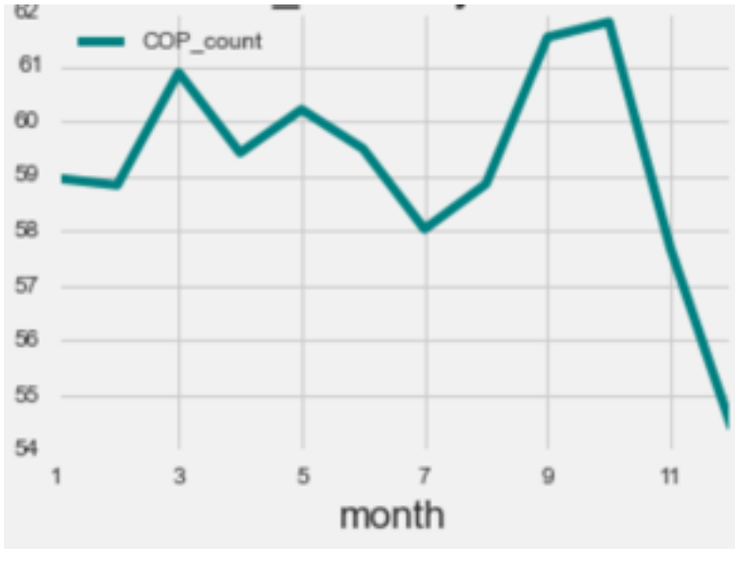
INITIAL OBSERVATIONS

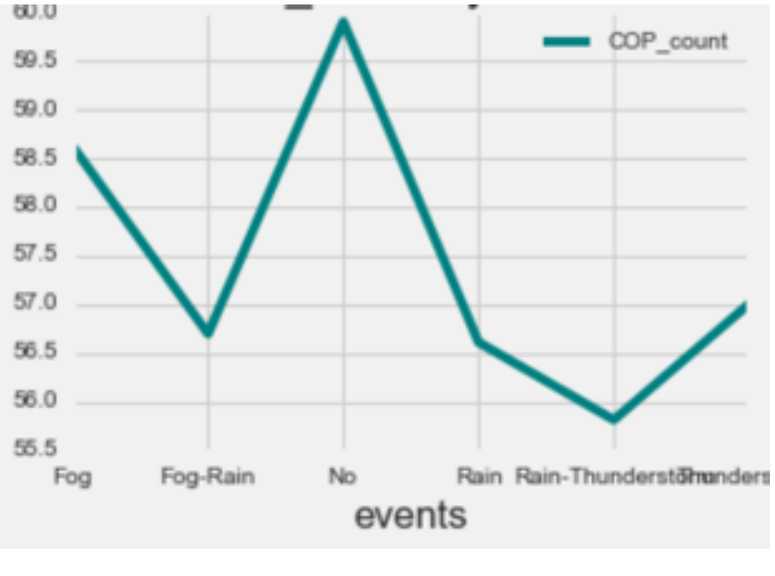






CATEGORICALS





CORRELATIONS

	temp_max	temp_mean	temp_delta	temp_apparent	dew_point_mean	humidity_mean	wind_chill_mean	heat_index
temp_max	1.000000	0.950802	0.734774	0.852066	0.746200	-0.440581	0.950802	0.696387
temp_mean	0.950802	1.000000	0.492378	0.930171	0.872325	-0.294023	1.000000	0.826268
temp_delta	0.734774	0.492378	1.000000	0.360566	0.185097	-0.618694	0.492378	0.144109
temp_apparent	0.852066	0.930171	0.360566	1.000000	0.957028	0.013004	0.930171	0.943559
dew_point_mean	0.746200	0.872325	0.185097	0.957028	1.000000	0.205574	0.872325	0.982994
humidity_mean	-0.440581	-0.294023	-0.618694	0.013004	0.205574	1.000000	-0.294023	0.257262
wind_chill_mean	0.950802	1.000000	0.492378	0.930171	0.872325	-0.294023	1.000000	0.826268
heat_index	0.696387	0.826268	0.144109	0.943559	0.982994	0.257262	0.826268	1.000000

- High level of Correlation between Features
- Engineered Feature calculated from them.
- Halo Effect of just more Variance

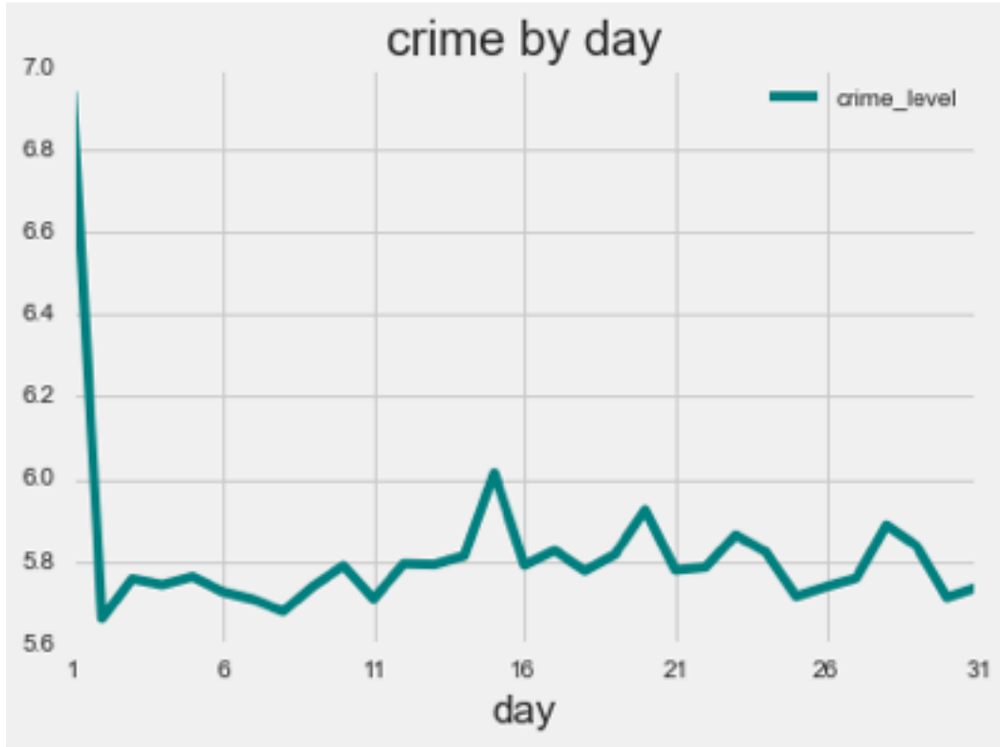
ISSUES

- Day One Issue

Abnormal number of crimes on day one

Many recorded at Time 00:01

Removed all Time 00:01 records



MODEL – CONCERNS

- Correlated data - Many of the features were based on temperature and humidity.
- Outliers - I was not sure if all the data was entered correctly.
- Complex/Nonlinear - Humidity can make the hot feel hotter and the cold seem colder.
- Ease of interoperability - Needed Feature Importance information

Choice - SVM or Random Forests, Linear Reg for back up and coefficients

FEATURES

MODELING RESULTS

	mean_abs_error	R2 score	time	lc_com
dummy	0.784282	-0.000157419	61.39	NaN
RF	0.774239	0.0257063	22676.5	.....
SVM	0.750965	0.0731648	66745.8	.....
LR	0.751183	0.0763004	52.1146	.....
knn	0.790226	-0.00414537	56354.1	.....

PREDICTIONS

	dayofweek	month	day	year	temp_max	prediction (dummy)	prediction (RF)	prediction (SVM)	prediction (LR)
0	monday	3	7	2016	58	60.672813	60.471429	58.400008	57.383723
1	tuesday	3	8	2016	58	60.672813	61.828571	59.679774	59.104755
2	wednesday	3	9	2016	64	60.672813	56.100000	61.300551	58.910556
3	thursday	3	10	2016	63	60.672813	59.450000	56.683531	55.597363
4	friday	3	11	2016	59	60.672813	56.364286	58.667685	56.911245
5	saturday	3	12	2016	59	60.672813	56.285714	58.863779	59.037412
6	sunday	3	13	2016	59	60.672813	57.907143	60.640057	60.806808
7	monday	3	14	2016	59	60.672813	59.450000	57.589806	56.884640
8	tuesday	3	15	2016	58	60.672813	58.314286	54.141668	55.016495
9	wednesday	3	16	2016	65	60.672813	57.914286	55.315704	54.135323

- MAE of ~8.5 makes these numbers highly suspect

CHALLENGES AND UNKNOWNS

- Changes in Policing Policies
  - Broken Windows - Stop and Frisk - New Technology
- Other Factors that effect Crime
  - Economic - Political - Education
- Using San Francisco as Test City
  - Generally Mild Weather conditions
- The complexity of Human Emotions

FUTURE POSSIBILITIES

- Focus more on weather feature
- Cumulative effects
- New Test City
- Refine the data and crime variables
  - Further refine the type of crime
  - Precinct in the city
  - Type of Officer needed (Street Cop, Detective, Desk Officer)
  - Time of Day and Shift
- Expand project with other factors

CONCLUSIONS

- Hard to get definitive Crime Prediction from the weather.
- Can get a general feel for higher or lower.
- Some correlation with Temperature, Rain and Clouds
- Probably no better than intuition of an experienced police officer.
- Crime and Human emotions are very complex.
- Did not find that the moon brought all the lunatics out.