7-FINAL

November 26, 2018

1 Assignment is below.

```
In [8]: from sklearn.svm import LinearSVC
        from sklearn.linear_model import LogisticRegression
        import pandas as pd
        import seaborn as sns
        import matplotlib.pyplot as plt
        %matplotlib inline
        sns.set(font_scale=1.5)
        import numpy as np

        from pylab import rcParams
        rcParams['figure.figsize'] = 20, 10


        from sklearn.linear_model import LogisticRegression as Model
```

Read in the Kobe Bryant shooting data [https://www.kaggle.com/c/kobe-bryant-shot-selection]

```
In [11]: kobe = pd.read_csv('../data/kobe.csv')
         kobe.dropna(inplace=True)

In [10]: list(kobe.columns)

Out[10]: ['action_type',
          'combined_shot_type',
          'game_event_id',
          'game_id',
          'lat',
          'loc_x',
          'loc_y',
          'lon',
          'minutes_remaining',
          'period',
          'playoffs',
          'season',
          'seconds_remaining',
```

```
 'shot_distance',
 'shot_made_flag',
 'shot_type',
 'shot_zone_area',
 'shot_zone_basic',
 'shot_zone_range',
 'team_id',
 'team_name',
 'game_date',
 'matchup',
 'opponent',
 'shot_id']
```

For now, use just the numerical datatypes. They are below as `num_columns`

```
In [12]: kobe.shot_zone_area.value_counts()

Out[12]: Center(C)                11289
         Right Side Center(RC)     3981
         Right Side(R)             3859
         Left Side Center(LC)      3364
         Left Side(L)              3132
         Back Court(BC)              72
         Name: shot_zone_area, dtype: int64

In [13]: kobe.shot_zone_range.value_counts()

Out[13]: Less Than 8 ft.    7857
         16-24 ft.          6907
         8-16 ft.           5580
         24+ ft.            5281
         Back Court Shot      72
         Name: shot_zone_range, dtype: int64

In [14]: kobe.shot_zone_basic.value_counts()

Out[14]: Mid-Range                10532
         Restricted Area           5932
         Above the Break 3         4720
         In The Paint (Non-RA)     3880
         Right Corner 3             333
         Left Corner 3              240
         Backcourt                   60
         Name: shot_zone_basic, dtype: int64

In [15]: kobe

Out[15]:              action_type combined_shot_type  game_event_id   game_id  \
         1               Jump Shot          Jump Shot             12  20000012
```

2	Jump Shot	Jump Shot	35	20000012
3	Jump Shot	Jump Shot	43	20000012
4	Driving Dunk Shot	Dunk	155	20000012
5	Jump Shot	Jump Shot	244	20000012
6	Layup Shot	Layup	251	20000012
8	Jump Shot	Jump Shot	265	20000012
9	Running Jump Shot	Jump Shot	294	20000012
10	Jump Shot	Jump Shot	309	20000012
11	Jump Shot	Jump Shot	4	20000019
12	Running Jump Shot	Jump Shot	27	20000019
13	Jump Shot	Jump Shot	66	20000019
14	Jump Shot	Jump Shot	80	20000019
15	Jump Shot	Jump Shot	86	20000019
17	Jump Shot	Jump Shot	138	20000019
18	Jump Shot	Jump Shot	244	20000019
20	Jump Shot	Jump Shot	255	20000019
21	Jump Shot	Jump Shot	265	20000019
22	Running Jump Shot	Jump Shot	274	20000019
23	Running Jump Shot	Jump Shot	299	20000019
24	Running Jump Shot	Jump Shot	307	20000019
25	Layup Shot	Layup	332	20000019
26	Jump Shot	Jump Shot	345	20000019
27	Jump Shot	Jump Shot	369	20000019
28	Jump Shot	Jump Shot	400	20000019
29	Jump Shot	Jump Shot	429	20000019
30	Running Jump Shot	Jump Shot	488	20000019
31	Jump Shot	Jump Shot	499	20000019
38	Jump Shot	Jump Shot	184	20000047
39	Jump Shot	Jump Shot	202	20000047
...	...	...	...	...
30661	Slam Dunk Shot	Dunk	245	49900087
30662	Jump Shot	Jump Shot	259	49900087
30663	Jump Shot	Jump Shot	270	49900087
30665	Layup Shot	Layup	280	49900087
30666	Jump Shot	Jump Shot	295	49900087
30667	Jump Shot	Jump Shot	368	49900087
30669	Jump Shot	Jump Shot	425	49900087
30670	Running Jump Shot	Jump Shot	15	49900088
30671	Driving Layup Shot	Layup	25	49900088
30672	Jump Shot	Jump Shot	29	49900088
30673	Jump Shot	Jump Shot	36	49900088
30674	Jump Shot	Jump Shot	81	49900088
30675	Jump Shot	Jump Shot	84	49900088
30676	Running Jump Shot	Jump Shot	98	49900088
30677	Jump Shot	Jump Shot	101	49900088
30678	Driving Layup Shot	Layup	181	49900088
30679	Layup Shot	Layup	212	49900088
30681	Jump Shot	Jump Shot	218	49900088

30683	Jump Shot	Jump Shot	228	49900088				
30684	Jump Shot	Jump Shot	231	49900088				
30685	Jump Shot	Jump Shot	249	49900088				
30687	Jump Shot	Jump Shot	284	49900088				
30688	Jump Shot	Jump Shot	308	49900088				
30689	Jump Shot	Jump Shot	326	49900088				
30690	Jump Shot	Jump Shot	331	49900088				
30691	Driving Layup Shot	Layup	382	49900088				
30692	Jump Shot	Jump Shot	397	49900088				
30694	Running Jump Shot	Jump Shot	426	49900088				
30695	Jump Shot	Jump Shot	448	49900088				
30696	Jump Shot	Jump Shot	471	49900088				

	lat	loc_x	loc_y	lon	minutes_remaining	period	...	\
1	34.0443	-157	0	-118.4268	10	1	...	
2	33.9093	-101	135	-118.3708	7	1	...	
3	33.8693	138	175	-118.1318	6	1	...	
4	34.0443	0	0	-118.2698	6	2	...	
5	34.0553	-145	-11	-118.4148	9	3	...	
6	34.0443	0	0	-118.2698	8	3	...	
8	33.9363	-65	108	-118.3348	6	3	...	
9	33.9193	-33	125	-118.3028	3	3	...	
10	33.8063	-94	238	-118.3638	1	3	...	
11	33.9173	121	127	-118.1488	11	1	...	
12	33.9343	-67	110	-118.3368	7	1	...	
13	34.0403	-94	4	-118.3638	2	1	...	
14	33.9973	-23	47	-118.2928	1	1	...	
15	33.8523	62	192	-118.2078	0	1	...	
17	33.8183	-117	226	-118.3868	8	2	...	
18	33.9473	-132	97	-118.4018	11	3	...	
20	33.9003	3	144	-118.2668	10	3	...	
21	33.9173	134	127	-118.1358	9	3	...	
22	33.9343	-16	110	-118.2858	7	3	...	
23	33.8943	-109	150	-118.3788	5	3	...	
24	33.9813	-46	63	-118.3158	5	3	...	
25	34.0443	0	0	-118.2698	2	3	...	
26	33.8483	-58	196	-118.3278	2	3	...	
27	33.8583	-183	186	-118.4528	0	3	...	
28	33.8713	85	173	-118.1848	8	4	...	
29	33.9573	3	87	-118.2668	6	4	...	
30	34.0403	121	4	-118.1488	1	4	...	
31	34.0103	127	34	-118.1428	0	4	...	
38	33.8603	91	184	-118.1788	3	2	...	
39	33.7723	-27	272	-118.2968	0	2	...	
...	...	...	...	...	...	...	...	
30661	34.0443	0	0	-118.2698	9	3	...	
30662	33.9913	-8	53	-118.2778	8	3	...	
30663	34.0193	106	25	-118.1638	6	3	...	

```
30665  34.0263    -14     18 -118.2838                5       3   ...
30666  33.8733    -81    171 -118.3508                4       3   ...
30667  33.7943     40    250 -118.2298                9       4   ...
30669  33.9913    171     53 -118.0988                3       4   ...
30670  34.0283    -74     16 -118.3438                9       1   ...
30671  34.0443      0      0 -118.2698                8       1   ...
30672  33.9893     89     55 -118.1808                8       1   ...
30673  34.0443    117      0 -118.1528                7       1   ...
30674  33.8283    117    216 -118.1528                2       1   ...
30675  33.8283   -134    216 -118.4038                2       1   ...
30676  34.0443   -141      0 -118.4108                0       1   ...
30677  33.9013   -113    143 -118.3828                0       1   ...
30678  34.0283     14     16 -118.2558                3       2   ...
30679  34.0443      0      0 -118.2698                0       2   ...
30681  33.7833    -18    261 -118.2878                0       2   ...
30683  33.8283      1    216 -118.2688               10       3   ...
30684  33.9553    -96     89 -118.3658               10       3   ...
30685  33.7943     81    250 -118.1888                7       3   ...
30687  33.9443     40    100 -118.2298                3       3   ...
30688  33.9833   -126     61 -118.3958                1       3   ...
30689  33.3653    -12    679 -118.2818                0       3   ...
30690  33.9443   -113    100 -118.3828               11       4   ...
30691  34.0443      0      0 -118.2698                7       4   ...
30692  33.9963      1     48 -118.2688                6       4   ...
30694  33.8783   -134    166 -118.4038                3       4   ...
30695  33.7773     31    267 -118.2388                2       4   ...
30696  33.9723      1     72 -118.2688                0       4   ...

            shot_type         shot_zone_area        shot_zone_basic  \
1      2PT Field Goal           Left Side(L)             Mid-Range
2      2PT Field Goal   Left Side Center(LC)             Mid-Range
3      2PT Field Goal  Right Side Center(RC)             Mid-Range
4      2PT Field Goal              Center(C)       Restricted Area
5      2PT Field Goal           Left Side(L)             Mid-Range
6      2PT Field Goal              Center(C)       Restricted Area
8      2PT Field Goal           Left Side(L)  In The Paint (Non-RA)
9      2PT Field Goal              Center(C)  In The Paint (Non-RA)
10     3PT Field Goal   Left Side Center(LC)     Above the Break 3
11     2PT Field Goal  Right Side Center(RC)             Mid-Range
12     2PT Field Goal           Left Side(L)  In The Paint (Non-RA)
13     2PT Field Goal           Left Side(L)             Mid-Range
14     2PT Field Goal              Center(C)  In The Paint (Non-RA)
15     2PT Field Goal              Center(C)             Mid-Range
17     3PT Field Goal   Left Side Center(LC)     Above the Break 3
18     2PT Field Goal   Left Side Center(LC)             Mid-Range
20     2PT Field Goal              Center(C)             Mid-Range
21     2PT Field Goal  Right Side Center(RC)             Mid-Range
22     2PT Field Goal              Center(C)  In The Paint (Non-RA)
```

```
23     2PT Field Goal   Left Side Center(LC)             Mid-Range
24     2PT Field Goal             Center(C)  In The Paint (Non-RA)
25     2PT Field Goal             Center(C)       Restricted Area
26     2PT Field Goal             Center(C)             Mid-Range
27     3PT Field Goal   Left Side Center(LC)     Above the Break 3
28     2PT Field Goal  Right Side Center(RC)             Mid-Range
29     2PT Field Goal             Center(C)  In The Paint (Non-RA)
30     2PT Field Goal          Right Side(R)             Mid-Range
31     2PT Field Goal          Right Side(R)             Mid-Range
38     2PT Field Goal  Right Side Center(RC)             Mid-Range
39     3PT Field Goal             Center(C)     Above the Break 3
...               ...                   ...                   ...
30661  2PT Field Goal             Center(C)       Restricted Area
30662  2PT Field Goal             Center(C)  In The Paint (Non-RA)
30663  2PT Field Goal          Right Side(R)             Mid-Range
30665  2PT Field Goal             Center(C)       Restricted Area
30666  2PT Field Goal   Left Side Center(LC)             Mid-Range
30667  3PT Field Goal             Center(C)     Above the Break 3
30669  2PT Field Goal          Right Side(R)             Mid-Range
30670  2PT Field Goal             Center(C)  In The Paint (Non-RA)
30671  2PT Field Goal             Center(C)       Restricted Area
30672  2PT Field Goal          Right Side(R)             Mid-Range
30673  2PT Field Goal          Right Side(R)             Mid-Range
30674  3PT Field Goal  Right Side Center(RC)     Above the Break 3
30675  3PT Field Goal   Left Side Center(LC)     Above the Break 3
30676  2PT Field Goal           Left Side(L)             Mid-Range
30677  2PT Field Goal   Left Side Center(LC)             Mid-Range
30678  2PT Field Goal             Center(C)       Restricted Area
30679  2PT Field Goal             Center(C)       Restricted Area
30681  3PT Field Goal             Center(C)     Above the Break 3
30683  2PT Field Goal             Center(C)             Mid-Range
30684  2PT Field Goal           Left Side(L)             Mid-Range
30685  3PT Field Goal             Center(C)     Above the Break 3
30687  2PT Field Goal             Center(C)  In The Paint (Non-RA)
30688  2PT Field Goal           Left Side(L)             Mid-Range
30689  3PT Field Goal         Back Court(BC)             Backcourt
30690  2PT Field Goal           Left Side(L)             Mid-Range
30691  2PT Field Goal             Center(C)       Restricted Area
30692  2PT Field Goal             Center(C)  In The Paint (Non-RA)
30694  2PT Field Goal   Left Side Center(LC)             Mid-Range
30695  3PT Field Goal             Center(C)     Above the Break 3
30696  2PT Field Goal             Center(C)  In The Paint (Non-RA)

       shot_zone_range     team_id           team_name   game_date  \
1            8-16 ft.  1610612747  Los Angeles Lakers  2000-10-31
2           16-24 ft.  1610612747  Los Angeles Lakers  2000-10-31
3           16-24 ft.  1610612747  Los Angeles Lakers  2000-10-31
4     Less Than 8 ft.  1610612747  Los Angeles Lakers  2000-10-31
```

5	8-16 ft.	1610612747	Los Angeles Lakers	2000-10-31
6	Less Than 8 ft.	1610612747	Los Angeles Lakers	2000-10-31
8	8-16 ft.	1610612747	Los Angeles Lakers	2000-10-31
9	8-16 ft.	1610612747	Los Angeles Lakers	2000-10-31
10	24+ ft.	1610612747	Los Angeles Lakers	2000-10-31
11	16-24 ft.	1610612747	Los Angeles Lakers	2000-11-01
12	8-16 ft.	1610612747	Los Angeles Lakers	2000-11-01
13	8-16 ft.	1610612747	Los Angeles Lakers	2000-11-01
14	Less Than 8 ft.	1610612747	Los Angeles Lakers	2000-11-01
15	16-24 ft.	1610612747	Los Angeles Lakers	2000-11-01
17	24+ ft.	1610612747	Los Angeles Lakers	2000-11-01
18	16-24 ft.	1610612747	Los Angeles Lakers	2000-11-01
20	8-16 ft.	1610612747	Los Angeles Lakers	2000-11-01
21	16-24 ft.	1610612747	Los Angeles Lakers	2000-11-01
22	8-16 ft.	1610612747	Los Angeles Lakers	2000-11-01
23	16-24 ft.	1610612747	Los Angeles Lakers	2000-11-01
24	Less Than 8 ft.	1610612747	Los Angeles Lakers	2000-11-01
25	Less Than 8 ft.	1610612747	Los Angeles Lakers	2000-11-01
26	16-24 ft.	1610612747	Los Angeles Lakers	2000-11-01
27	24+ ft.	1610612747	Los Angeles Lakers	2000-11-01
28	16-24 ft.	1610612747	Los Angeles Lakers	2000-11-01
29	8-16 ft.	1610612747	Los Angeles Lakers	2000-11-01
30	8-16 ft.	1610612747	Los Angeles Lakers	2000-11-01
31	8-16 ft.	1610612747	Los Angeles Lakers	2000-11-01
38	16-24 ft.	1610612747	Los Angeles Lakers	2000-11-04
39	24+ ft.	1610612747	Los Angeles Lakers	2000-11-04
...	...	...	...	...
30661	Less Than 8 ft.	1610612747	Los Angeles Lakers	2000-06-16
30662	Less Than 8 ft.	1610612747	Los Angeles Lakers	2000-06-16
30663	8-16 ft.	1610612747	Los Angeles Lakers	2000-06-16
30665	Less Than 8 ft.	1610612747	Los Angeles Lakers	2000-06-16
30666	16-24 ft.	1610612747	Los Angeles Lakers	2000-06-16
30667	24+ ft.	1610612747	Los Angeles Lakers	2000-06-16
30669	16-24 ft.	1610612747	Los Angeles Lakers	2000-06-16
30670	Less Than 8 ft.	1610612747	Los Angeles Lakers	2000-06-19
30671	Less Than 8 ft.	1610612747	Los Angeles Lakers	2000-06-19
30672	8-16 ft.	1610612747	Los Angeles Lakers	2000-06-19
30673	8-16 ft.	1610612747	Los Angeles Lakers	2000-06-19
30674	24+ ft.	1610612747	Los Angeles Lakers	2000-06-19
30675	24+ ft.	1610612747	Los Angeles Lakers	2000-06-19
30676	8-16 ft.	1610612747	Los Angeles Lakers	2000-06-19
30677	16-24 ft.	1610612747	Los Angeles Lakers	2000-06-19
30678	Less Than 8 ft.	1610612747	Los Angeles Lakers	2000-06-19
30679	Less Than 8 ft.	1610612747	Los Angeles Lakers	2000-06-19
30681	24+ ft.	1610612747	Los Angeles Lakers	2000-06-19
30683	16-24 ft.	1610612747	Los Angeles Lakers	2000-06-19
30684	8-16 ft.	1610612747	Los Angeles Lakers	2000-06-19
30685	24+ ft.	1610612747	Los Angeles Lakers	2000-06-19

```
30687         8-16 ft.  1610612747  Los Angeles Lakers  2000-06-19
30688         8-16 ft.  1610612747  Los Angeles Lakers  2000-06-19
30689  Back Court Shot  1610612747  Los Angeles Lakers  2000-06-19
30690         8-16 ft.  1610612747  Los Angeles Lakers  2000-06-19
30691  Less Than 8 ft.  1610612747  Los Angeles Lakers  2000-06-19
30692  Less Than 8 ft.  1610612747  Los Angeles Lakers  2000-06-19
30694        16-24 ft.  1610612747  Los Angeles Lakers  2000-06-19
30695          24+ ft.  1610612747  Los Angeles Lakers  2000-06-19
30696  Less Than 8 ft.  1610612747  Los Angeles Lakers  2000-06-19
```

	matchup	opponent	shot_id
1	LAL @ POR	POR	2
2	LAL @ POR	POR	3
3	LAL @ POR	POR	4
4	LAL @ POR	POR	5
5	LAL @ POR	POR	6
6	LAL @ POR	POR	7
8	LAL @ POR	POR	9
9	LAL @ POR	POR	10
10	LAL @ POR	POR	11
11	LAL vs. UTA	UTA	12
12	LAL vs. UTA	UTA	13
13	LAL vs. UTA	UTA	14
14	LAL vs. UTA	UTA	15
15	LAL vs. UTA	UTA	16
17	LAL vs. UTA	UTA	18
18	LAL vs. UTA	UTA	19
20	LAL vs. UTA	UTA	21
21	LAL vs. UTA	UTA	22
22	LAL vs. UTA	UTA	23
23	LAL vs. UTA	UTA	24
24	LAL vs. UTA	UTA	25
25	LAL vs. UTA	UTA	26
26	LAL vs. UTA	UTA	27
27	LAL vs. UTA	UTA	28
28	LAL vs. UTA	UTA	29
29	LAL vs. UTA	UTA	30
30	LAL vs. UTA	UTA	31
31	LAL vs. UTA	UTA	32
38	LAL @ VAN	VAN	39
39	LAL @ VAN	VAN	40
...	...	...	...
30661	LAL @ IND	IND	30662
30662	LAL @ IND	IND	30663
30663	LAL @ IND	IND	30664
30665	LAL @ IND	IND	30666
30666	LAL @ IND	IND	30667
30667	LAL @ IND	IND	30668

```
30669    LAL @ IND      IND    30670
30670  LAL vs. IND      IND    30671
30671  LAL vs. IND      IND    30672
30672  LAL vs. IND      IND    30673
30673  LAL vs. IND      IND    30674
30674  LAL vs. IND      IND    30675
30675  LAL vs. IND      IND    30676
30676  LAL vs. IND      IND    30677
30677  LAL vs. IND      IND    30678
30678  LAL vs. IND      IND    30679
30679  LAL vs. IND      IND    30680
30681  LAL vs. IND      IND    30682
30683  LAL vs. IND      IND    30684
30684  LAL vs. IND      IND    30685
30685  LAL vs. IND      IND    30686
30687  LAL vs. IND      IND    30688
30688  LAL vs. IND      IND    30689
30689  LAL vs. IND      IND    30690
30690  LAL vs. IND      IND    30691
30691  LAL vs. IND      IND    30692
30692  LAL vs. IND      IND    30693
30694  LAL vs. IND      IND    30695
30695  LAL vs. IND      IND    30696
30696  LAL vs. IND      IND    30697

[25697 rows x 25 columns]
```

In [16]:
```
kobe.shot_made_flag.value_counts(normalize=True)
```

Out[16]:
```
0.0    0.553839
1.0    0.446161
Name: shot_made_flag, dtype: float64
```

In [17]:
```
kobe.shot_made_flag.value_counts(normalize=False)
```

Out[17]:
```
0.0    14232
1.0    11465
Name: shot_made_flag, dtype: int64
```

In [18]:
```
num_columns = [col for col, dtype in zip(kobe.columns, kobe.dtypes) if dtype != 'obje
num_columns
```

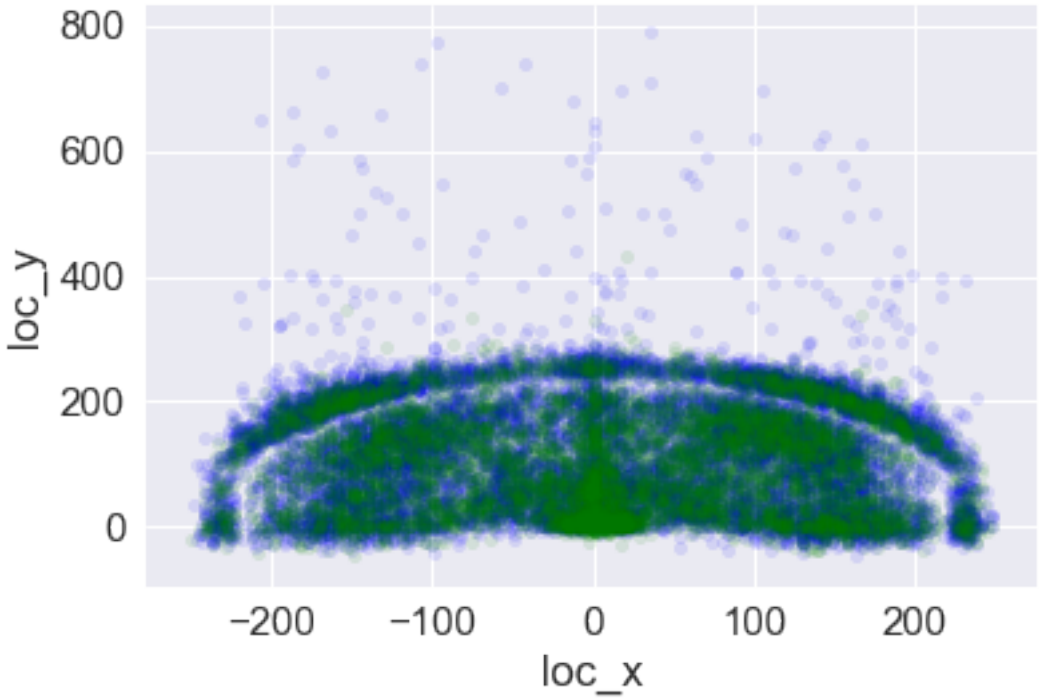
Out[18]:
```
['game_event_id',
 'game_id',
 'lat',
 'loc_x',
 'loc_y',
 'lon',
 'minutes_remaining',
```

```
'period',
'playoffs',
'seconds_remaining',
'shot_distance',
'shot_made_flag',
'team_id',
'shot_id']
```

1.1 The `shot_made_flag` is the result (0 or 1) of the shot that Kobe took. Some of the values are missing (e.g. `NaN`) but we dropped them.

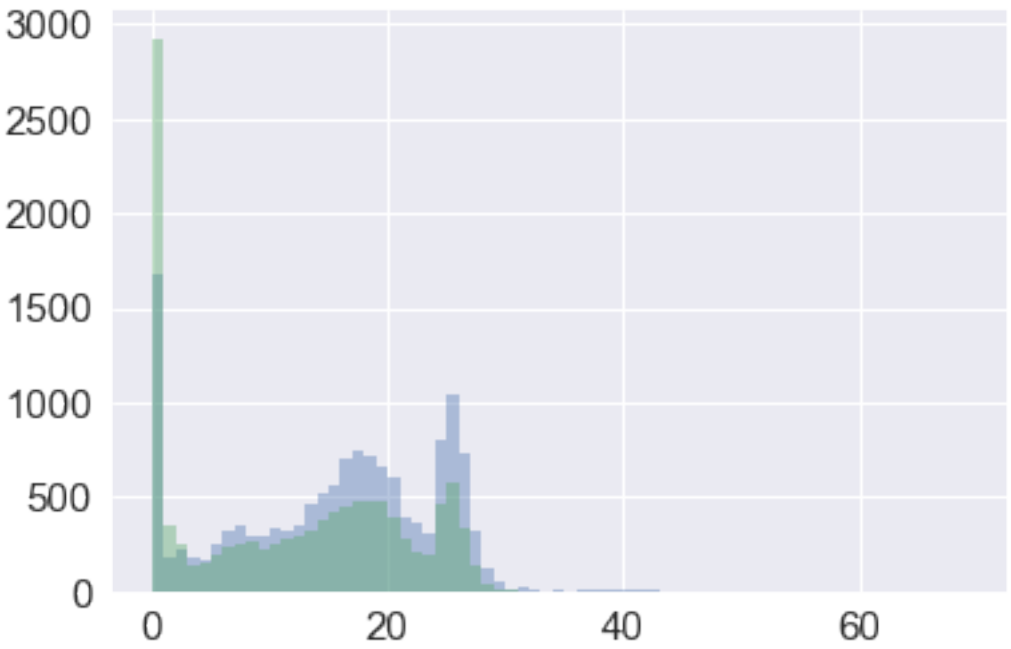
In [19]:
```
fig, ax = plt.subplots()
kobe[kobe.shot_made_flag==0].plot(kind='scatter', x='loc_x', y='loc_y', color='blue',
kobe[kobe.shot_made_flag==1].plot(kind='scatter', x='loc_x', y='loc_y', color='green'
# plt.scatter(kobe.loc_x, kobe.loc_y, alpha=0.2)
```

Out[19]: `<matplotlib.axes._subplots.AxesSubplot at 0x1a9654b2ba8>`



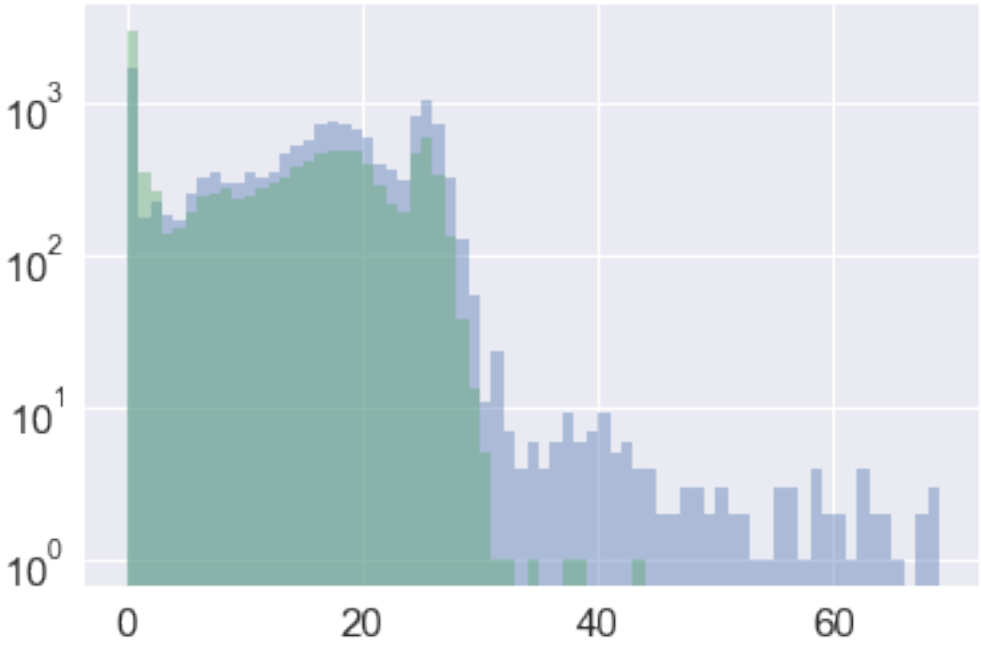
In [20]:
```
kobe[kobe.shot_made_flag==0].shot_distance.hist(bins=np.arange(0,70,1), alpha=.4)
kobe[kobe.shot_made_flag==1].shot_distance.hist(bins=np.arange(0,70,1), alpha=.4)
```

Out[20]: `<matplotlib.axes._subplots.AxesSubplot at 0x1a96563aef0>`



In [21]:
```
kobe[kobe.shot_made_flag==0].shot_distance.hist(bins=np.arange(0,70,1), alpha=.4, log=
kobe[kobe.shot_made_flag==1].shot_distance.hist(bins=np.arange(0,70,1), alpha=.4, log=
```

Out[21]: <matplotlib.axes._subplots.AxesSubplot at 0x1a9640bb470>

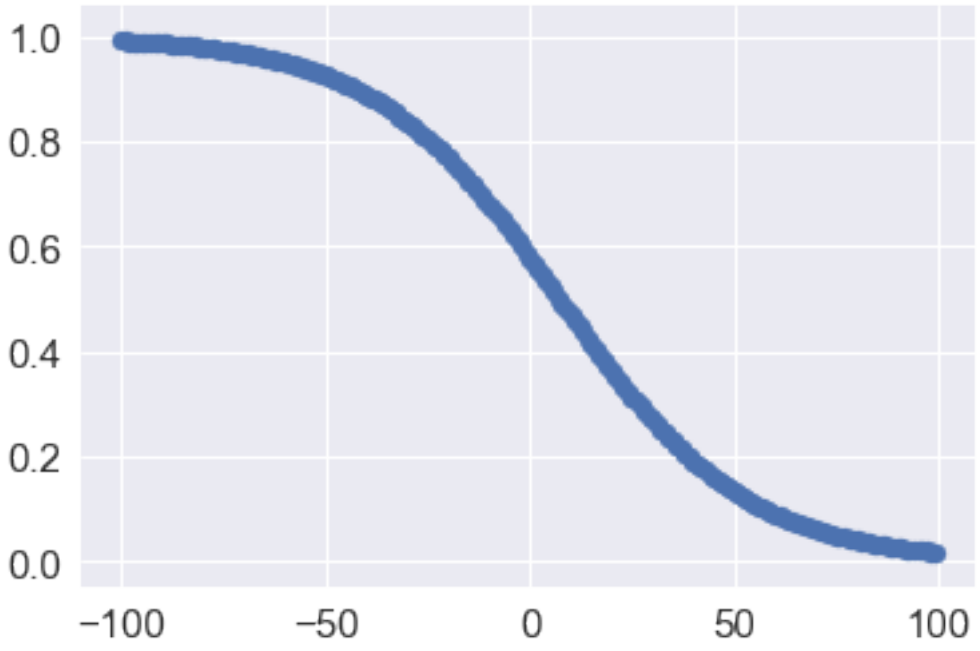


In [22]:
```
# fit a linear regression model and store the predictions
feature_cols = ['shot_distance', 'minutes_remaining']
X = kobe[feature_cols]
y = kobe.shot_made_flag

model = Model()
model.fit(X, y)
kobe['pred'] = model.predict(X)


from sklearn.metrics import accuracy_score
accuracy_score(kobe.shot_made_flag, kobe.pred.round())
```

Out[22]:
```
0.5971903335019653
```

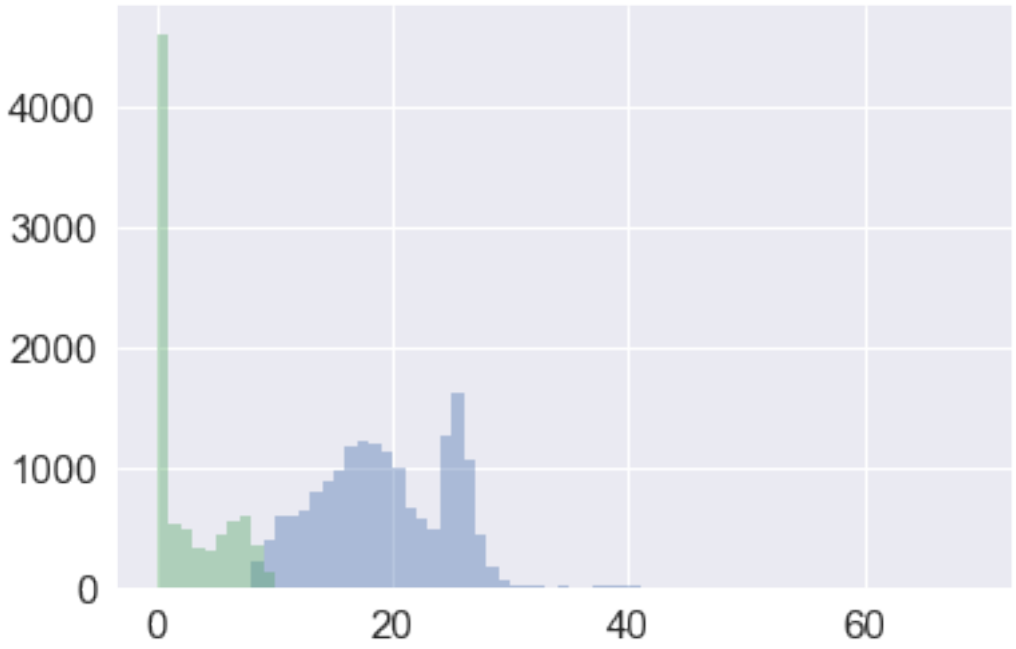
In [23]:
```
distances = np.arange(-100, 100)
minutes = np.array([0]*200)
x_trial = np.column_stack((distances, minutes))
model.predict_proba(x_trial)
plt.scatter(distances, model.predict_proba(x_trial)[:,1])
```

Out[23]:
```
<matplotlib.collections.PathCollection at 0x1a96610dac8>
```



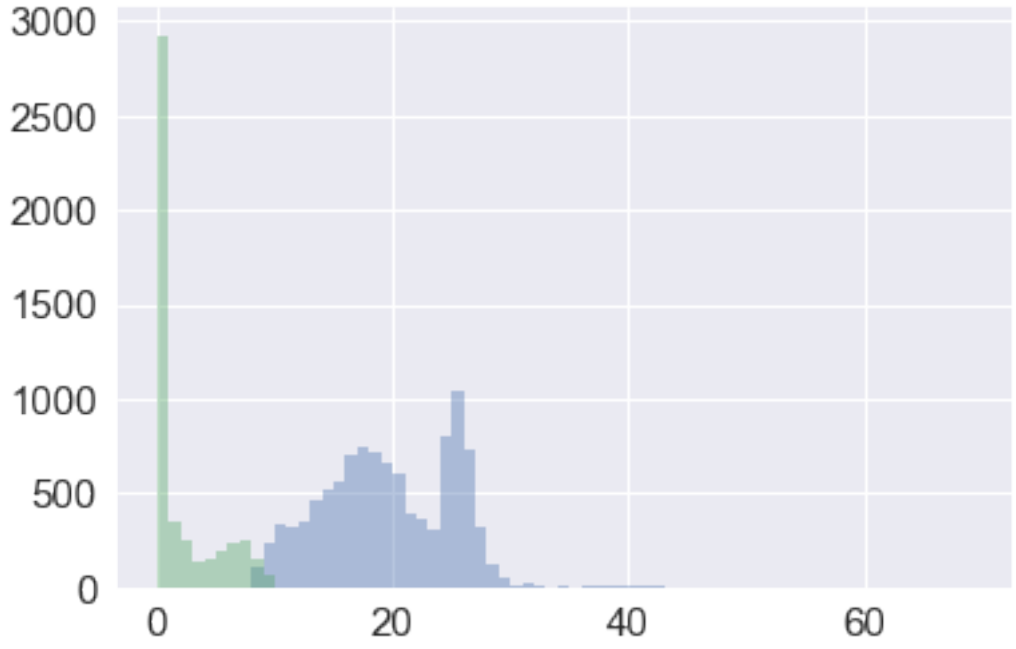
In [24]:
```
kobe[(kobe.pred==0)].shot_distance.hist(bins=np.arange(0,70,1), alpha=.4)
kobe[(kobe.pred==1)].shot_distance.hist(bins=np.arange(0,70,1), alpha=.4)
```

Out[24]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1a9660389b0>
```



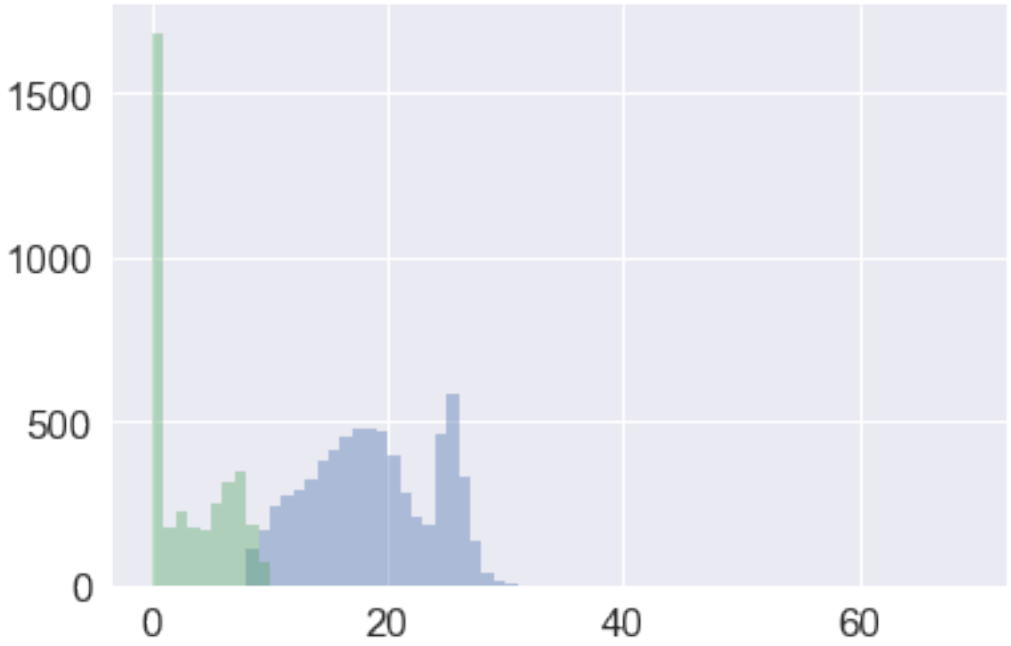
In [25]:
```
kobe[(kobe.pred==0) & (kobe.shot_made_flag==0)].shot_distance.hist(bins=np.arange(0,7
kobe[(kobe.pred==1) & (kobe.shot_made_flag==1)].shot_distance.hist(bins=np.arange(0,7
```

Out[25]: <matplotlib.axes._subplots.AxesSubplot at 0x1a9672205f8>



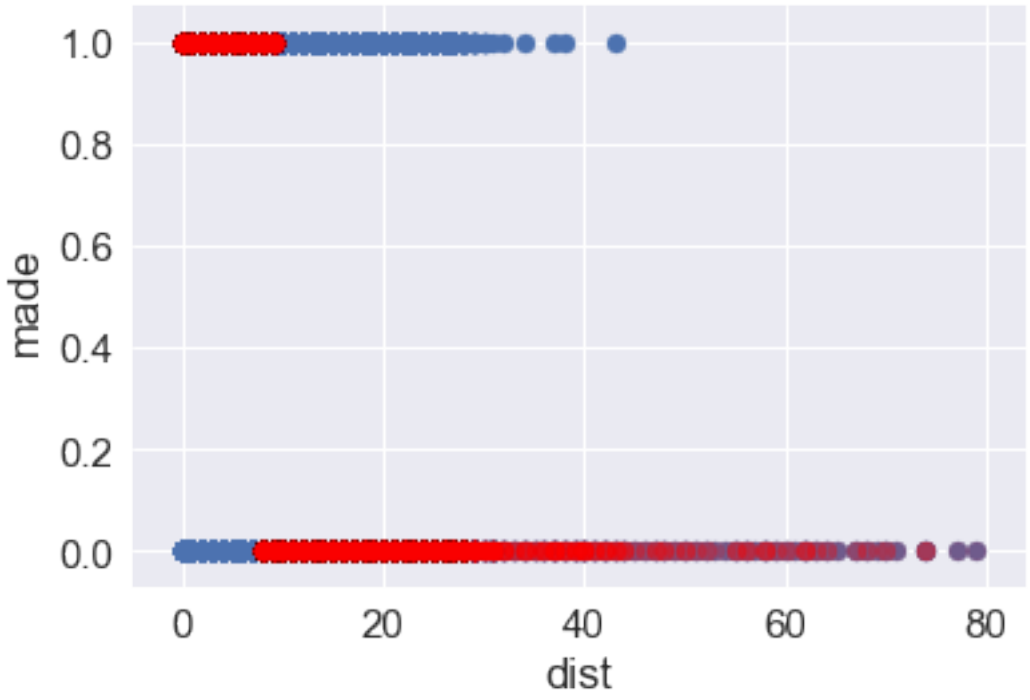
In [26]:
```
kobe[(kobe.pred==0) & (kobe.shot_made_flag==1)].shot_distance.hist(bins=np.arange(0,7
kobe[(kobe.pred==1) & (kobe.shot_made_flag==0)].shot_distance.hist(bins=np.arange(0,7
```

Out[26]:
```
<matplotlib.axes._subplots.AxesSubplot at 0x1a967418cc0>
```



In [27]:
```
# scatter plot that includes the regression line
plt.scatter(kobe.shot_distance, kobe.shot_made_flag)
plt.scatter(kobe.shot_distance, kobe.pred, color='red', alpha=.2)
plt.xlabel('dist')
plt.ylabel('made')
```

Out[27]:
```
Text(0,0.5,'made')
```



1.2 The following is a reminder of how the SciKit-Learn Models can be interfaced

```
In [28]: from sklearn.linear_model import LogisticRegression as Model
         # from sklearn.tree import DecisionTreeClassifier as Model
         # from sklearn.ensemble import RandomForestClassifier as Model
         model = Model()

         from sklearn.metrics import (accuracy_score,
                                      classification_report,
                                      confusion_matrix, auc, roc_curve
                                     )
         from sklearn.metrics import *
         from sklearn import cross_validation

         X_train, X_test, y_train, y_test = cross_validation.train_test_split(
             X, y, test_size=0.4, random_state=0)

         cross_validation.cross_val_score(model, X, y, cv=10)
```

```
C:\Users\Erin\Anaconda3\lib\site-packages\sklearn\cross_validation.py:41: DeprecationWarning:
  "This module will be removed in 0.20.", DeprecationWarning)
```

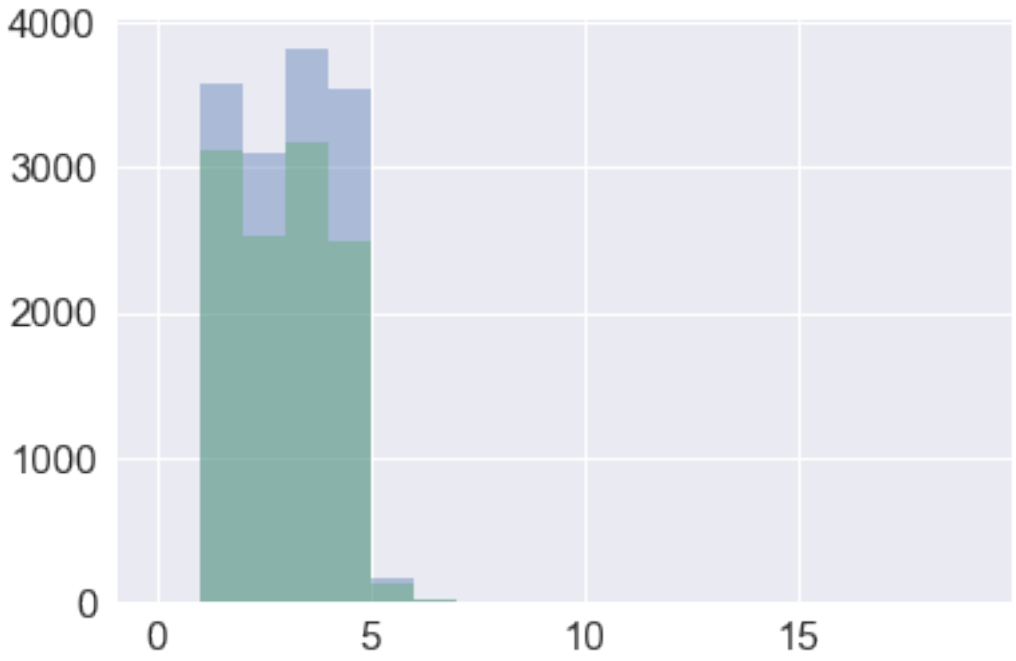
```
Out[28]: array([0.59237651, 0.59354337, 0.59299611, 0.59688716, 0.61750973,
                0.58388478, 0.60334761, 0.60918645, 0.60140132, 0.58388478])
```

2 Assignment

2.0.1 Warmup. Perform some analysis on Kobe's shot selection. Ask and answer (with charts) questions such as: Does Kobe make more shots in the 4th quarter than on average? Does Kobe make more shots from the left more than the right? What was Kobe's best year for shooting percentage? Etc. The more naunced the more you'll have a feel for the data.

In [29]:
```
kobe[kobe.shot_made_flag==0].period.hist(bins=np.arange(0,20,1), alpha=.4)
kobe[kobe.shot_made_flag==1].period.hist(bins=np.arange(0,20,1), alpha=.4)
```

Out[29]: <matplotlib.axes._subplots.AxesSubplot at 0x1a967702240>



2.0.2 1. Create a new column called `abs_x` that is equal to the absolute value of loc_x. Plot a histogram of made shots and missed shots using this variable. Explain in detail (with graphics and evidence) why this could be a better feature/column to use in a Logsitic Regression model instead of `loc_x`.

In [46]:
```
df = pd.DataFrame(data=kobe)
df
```

Out[46]:

	action_type	combined_shot_type	game_event_id	game_id	\
1	Jump Shot	Jump Shot	12	20000012	
2	Jump Shot	Jump Shot	35	20000012	
3	Jump Shot	Jump Shot	43	20000012	
4	Driving Dunk Shot	Dunk	155	20000012	
5	Jump Shot	Jump Shot	244	20000012	

6	Layup Shot	Layup	251	20000012
8	Jump Shot	Jump Shot	265	20000012
9	Running Jump Shot	Jump Shot	294	20000012
10	Jump Shot	Jump Shot	309	20000012
11	Jump Shot	Jump Shot	4	20000019
12	Running Jump Shot	Jump Shot	27	20000019
13	Jump Shot	Jump Shot	66	20000019
14	Jump Shot	Jump Shot	80	20000019
15	Jump Shot	Jump Shot	86	20000019
17	Jump Shot	Jump Shot	138	20000019
18	Jump Shot	Jump Shot	244	20000019
20	Jump Shot	Jump Shot	255	20000019
21	Jump Shot	Jump Shot	265	20000019
22	Running Jump Shot	Jump Shot	274	20000019
23	Running Jump Shot	Jump Shot	299	20000019
24	Running Jump Shot	Jump Shot	307	20000019
25	Layup Shot	Layup	332	20000019
26	Jump Shot	Jump Shot	345	20000019
27	Jump Shot	Jump Shot	369	20000019
28	Jump Shot	Jump Shot	400	20000019
29	Jump Shot	Jump Shot	429	20000019
30	Running Jump Shot	Jump Shot	488	20000019
31	Jump Shot	Jump Shot	499	20000019
38	Jump Shot	Jump Shot	184	20000047
39	Jump Shot	Jump Shot	202	20000047
...	...	...	...	...
30661	Slam Dunk Shot	Dunk	245	49900087
30662	Jump Shot	Jump Shot	259	49900087
30663	Jump Shot	Jump Shot	270	49900087
30665	Layup Shot	Layup	280	49900087
30666	Jump Shot	Jump Shot	295	49900087
30667	Jump Shot	Jump Shot	368	49900087
30669	Jump Shot	Jump Shot	425	49900087
30670	Running Jump Shot	Jump Shot	15	49900088
30671	Driving Layup Shot	Layup	25	49900088
30672	Jump Shot	Jump Shot	29	49900088
30673	Jump Shot	Jump Shot	36	49900088
30674	Jump Shot	Jump Shot	81	49900088
30675	Jump Shot	Jump Shot	84	49900088
30676	Running Jump Shot	Jump Shot	98	49900088
30677	Jump Shot	Jump Shot	101	49900088
30678	Driving Layup Shot	Layup	181	49900088
30679	Layup Shot	Layup	212	49900088
30681	Jump Shot	Jump Shot	218	49900088
30683	Jump Shot	Jump Shot	228	49900088
30684	Jump Shot	Jump Shot	231	49900088
30685	Jump Shot	Jump Shot	249	49900088
30687	Jump Shot	Jump Shot	284	49900088

```
30688           Jump Shot          Jump Shot             308  49900088
30689           Jump Shot          Jump Shot             326  49900088
30690           Jump Shot          Jump Shot             331  49900088
30691  Driving Layup Shot              Layup             382  49900088
30692           Jump Shot          Jump Shot             397  49900088
30694   Running Jump Shot          Jump Shot             426  49900088
30695           Jump Shot          Jump Shot             448  49900088
30696           Jump Shot          Jump Shot             471  49900088

           lat  loc_x  loc_y       lon  minutes_remaining  period  ...   \
1      34.0443   -157      0 -118.4268                 10       1  ...
2      33.9093   -101    135 -118.3708                  7       1  ...
3      33.8693    138    175 -118.1318                  6       1  ...
4      34.0443      0      0 -118.2698                  6       2  ...
5      34.0553   -145    -11 -118.4148                  9       3  ...
6      34.0443      0      0 -118.2698                  8       3  ...
8      33.9363    -65    108 -118.3348                  6       3  ...
9      33.9193    -33    125 -118.3028                  3       3  ...
10     33.8063    -94    238 -118.3638                  1       3  ...
11     33.9173    121    127 -118.1488                 11       1  ...
12     33.9343    -67    110 -118.3368                  7       1  ...
13     34.0403    -94      4 -118.3638                  2       1  ...
14     33.9973    -23     47 -118.2928                  1       1  ...
15     33.8523     62    192 -118.2078                  0       1  ...
17     33.8183   -117    226 -118.3868                  8       2  ...
18     33.9473   -132     97 -118.4018                 11       3  ...
20     33.9003      3    144 -118.2668                 10       3  ...
21     33.9173    134    127 -118.1358                  9       3  ...
22     33.9343    -16    110 -118.2858                  7       3  ...
23     33.8943   -109    150 -118.3788                  5       3  ...
24     33.9813    -46     63 -118.3158                  5       3  ...
25     34.0443      0      0 -118.2698                  2       3  ...
26     33.8483    -58    196 -118.3278                  2       3  ...
27     33.8583   -183    186 -118.4528                  0       3  ...
28     33.8713     85    173 -118.1848                  8       4  ...
29     33.9573      3     87 -118.2668                  6       4  ...
30     34.0403    121      4 -118.1488                  1       4  ...
31     34.0103    127     34 -118.1428                  0       4  ...
38     33.8603     91    184 -118.1788                  3       2  ...
39     33.7723    -27    272 -118.2968                  0       2  ...
...        ...    ...    ...       ...                ...     ...  ...
30661  34.0443      0      0 -118.2698                  9       3  ...
30662  33.9913     -8     53 -118.2778                  8       3  ...
30663  34.0193    106     25 -118.1638                  6       3  ...
30665  34.0263    -14     18 -118.2838                  5       3  ...
30666  33.8733    -81    171 -118.3508                  4       3  ...
30667  33.7943     40    250 -118.2298                  9       4  ...
30669  33.9913    171     53 -118.0988                  3       4  ...
```

```
30670  34.0283    -74     16 -118.3438                 9       1  ...
30671  34.0443      0      0 -118.2698                 8       1  ...
30672  33.9893     89     55 -118.1808                 8       1  ...
30673  34.0443    117      0 -118.1528                 7       1  ...
30674  33.8283    117    216 -118.1528                 2       1  ...
30675  33.8283   -134    216 -118.4038                 2       1  ...
30676  34.0443   -141      0 -118.4108                 0       1  ...
30677  33.9013   -113    143 -118.3828                 0       1  ...
30678  34.0283     14     16 -118.2558                 3       2  ...
30679  34.0443      0      0 -118.2698                 0       2  ...
30681  33.7833    -18    261 -118.2878                 0       2  ...
30683  33.8283      1    216 -118.2688                10       3  ...
30684  33.9553    -96     89 -118.3658                10       3  ...
30685  33.7943     81    250 -118.1888                 7       3  ...
30687  33.9443     40    100 -118.2298                 3       3  ...
30688  33.9833   -126     61 -118.3958                 1       3  ...
30689  33.3653    -12    679 -118.2818                 0       3  ...
30690  33.9443   -113    100 -118.3828                11       4  ...
30691  34.0443      0      0 -118.2698                 7       4  ...
30692  33.9963      1     48 -118.2688                 6       4  ...
30694  33.8783   -134    166 -118.4038                 3       4  ...
30695  33.7773     31    267 -118.2388                 2       4  ...
30696  33.9723      1     72 -118.2688                 0       4  ...

                shot_zone_area        shot_zone_basic  shot_zone_range  \
1                Left Side(L)              Mid-Range        8-16 ft.
2       Left Side Center(LC)              Mid-Range       16-24 ft.
3      Right Side Center(RC)              Mid-Range       16-24 ft.
4                   Center(C)       Restricted Area  Less Than 8 ft.
5                Left Side(L)              Mid-Range        8-16 ft.
6                   Center(C)       Restricted Area  Less Than 8 ft.
8                Left Side(L)  In The Paint (Non-RA)        8-16 ft.
9                   Center(C)  In The Paint (Non-RA)        8-16 ft.
10      Left Side Center(LC)      Above the Break 3         24+ ft.
11     Right Side Center(RC)              Mid-Range       16-24 ft.
12               Left Side(L)  In The Paint (Non-RA)        8-16 ft.
13               Left Side(L)              Mid-Range        8-16 ft.
14                  Center(C)  In The Paint (Non-RA)  Less Than 8 ft.
15                  Center(C)              Mid-Range       16-24 ft.
17      Left Side Center(LC)      Above the Break 3         24+ ft.
18      Left Side Center(LC)              Mid-Range       16-24 ft.
20                  Center(C)              Mid-Range        8-16 ft.
21     Right Side Center(RC)              Mid-Range       16-24 ft.
22                  Center(C)  In The Paint (Non-RA)        8-16 ft.
23      Left Side Center(LC)              Mid-Range       16-24 ft.
24                  Center(C)  In The Paint (Non-RA)  Less Than 8 ft.
25                  Center(C)       Restricted Area  Less Than 8 ft.
26                  Center(C)              Mid-Range       16-24 ft.
```

```
27      Left Side Center(LC)      Above the Break 3          24+ ft.
28     Right Side Center(RC)              Mid-Range        16-24 ft.
29                Center(C)  In The Paint (Non-RA)         8-16 ft.
30             Right Side(R)              Mid-Range         8-16 ft.
31             Right Side(R)              Mid-Range         8-16 ft.
38     Right Side Center(RC)              Mid-Range        16-24 ft.
39                Center(C)      Above the Break 3          24+ ft.
...                     ...                    ...             ...
30661             Center(C)        Restricted Area  Less Than 8 ft.
30662             Center(C)  In The Paint (Non-RA)  Less Than 8 ft.
30663          Right Side(R)              Mid-Range         8-16 ft.
30665             Center(C)        Restricted Area  Less Than 8 ft.
30666   Left Side Center(LC)              Mid-Range        16-24 ft.
30667             Center(C)      Above the Break 3          24+ ft.
30669          Right Side(R)              Mid-Range        16-24 ft.
30670             Center(C)  In The Paint (Non-RA)  Less Than 8 ft.
30671             Center(C)        Restricted Area  Less Than 8 ft.
30672          Right Side(R)              Mid-Range         8-16 ft.
30673          Right Side(R)              Mid-Range         8-16 ft.
30674  Right Side Center(RC)      Above the Break 3          24+ ft.
30675   Left Side Center(LC)      Above the Break 3          24+ ft.
30676           Left Side(L)              Mid-Range         8-16 ft.
30677   Left Side Center(LC)              Mid-Range        16-24 ft.
30678             Center(C)        Restricted Area  Less Than 8 ft.
30679             Center(C)        Restricted Area  Less Than 8 ft.
30681             Center(C)      Above the Break 3          24+ ft.
30683             Center(C)              Mid-Range        16-24 ft.
30684           Left Side(L)              Mid-Range         8-16 ft.
30685             Center(C)      Above the Break 3          24+ ft.
30687             Center(C)  In The Paint (Non-RA)         8-16 ft.
30688           Left Side(L)              Mid-Range         8-16 ft.
30689         Back Court(BC)               Backcourt  Back Court Shot
30690           Left Side(L)              Mid-Range         8-16 ft.
30691             Center(C)        Restricted Area  Less Than 8 ft.
30692             Center(C)  In The Paint (Non-RA)  Less Than 8 ft.
30694   Left Side Center(LC)              Mid-Range        16-24 ft.
30695             Center(C)      Above the Break 3          24+ ft.
30696             Center(C)  In The Paint (Non-RA)  Less Than 8 ft.

         team_id           team_name   game_date      matchup opponent  \
1     1610612747  Los Angeles Lakers  2000-10-31    LAL @ POR      POR
2     1610612747  Los Angeles Lakers  2000-10-31    LAL @ POR      POR
3     1610612747  Los Angeles Lakers  2000-10-31    LAL @ POR      POR
4     1610612747  Los Angeles Lakers  2000-10-31    LAL @ POR      POR
5     1610612747  Los Angeles Lakers  2000-10-31    LAL @ POR      POR
6     1610612747  Los Angeles Lakers  2000-10-31    LAL @ POR      POR
8     1610612747  Los Angeles Lakers  2000-10-31    LAL @ POR      POR
9     1610612747  Los Angeles Lakers  2000-10-31    LAL @ POR      POR
```

10	1610612747	Los Angeles Lakers	2000-10-31	LAL @ POR	POR
11	1610612747	Los Angeles Lakers	2000-11-01	LAL vs. UTA	UTA
12	1610612747	Los Angeles Lakers	2000-11-01	LAL vs. UTA	UTA
13	1610612747	Los Angeles Lakers	2000-11-01	LAL vs. UTA	UTA
14	1610612747	Los Angeles Lakers	2000-11-01	LAL vs. UTA	UTA
15	1610612747	Los Angeles Lakers	2000-11-01	LAL vs. UTA	UTA
17	1610612747	Los Angeles Lakers	2000-11-01	LAL vs. UTA	UTA
18	1610612747	Los Angeles Lakers	2000-11-01	LAL vs. UTA	UTA
20	1610612747	Los Angeles Lakers	2000-11-01	LAL vs. UTA	UTA
21	1610612747	Los Angeles Lakers	2000-11-01	LAL vs. UTA	UTA
22	1610612747	Los Angeles Lakers	2000-11-01	LAL vs. UTA	UTA
23	1610612747	Los Angeles Lakers	2000-11-01	LAL vs. UTA	UTA
24	1610612747	Los Angeles Lakers	2000-11-01	LAL vs. UTA	UTA
25	1610612747	Los Angeles Lakers	2000-11-01	LAL vs. UTA	UTA
26	1610612747	Los Angeles Lakers	2000-11-01	LAL vs. UTA	UTA
27	1610612747	Los Angeles Lakers	2000-11-01	LAL vs. UTA	UTA
28	1610612747	Los Angeles Lakers	2000-11-01	LAL vs. UTA	UTA
29	1610612747	Los Angeles Lakers	2000-11-01	LAL vs. UTA	UTA
30	1610612747	Los Angeles Lakers	2000-11-01	LAL vs. UTA	UTA
31	1610612747	Los Angeles Lakers	2000-11-01	LAL vs. UTA	UTA
38	1610612747	Los Angeles Lakers	2000-11-04	LAL @ VAN	VAN
39	1610612747	Los Angeles Lakers	2000-11-04	LAL @ VAN	VAN
...	...	...	...	...	...
30661	1610612747	Los Angeles Lakers	2000-06-16	LAL @ IND	IND
30662	1610612747	Los Angeles Lakers	2000-06-16	LAL @ IND	IND
30663	1610612747	Los Angeles Lakers	2000-06-16	LAL @ IND	IND
30665	1610612747	Los Angeles Lakers	2000-06-16	LAL @ IND	IND
30666	1610612747	Los Angeles Lakers	2000-06-16	LAL @ IND	IND
30667	1610612747	Los Angeles Lakers	2000-06-16	LAL @ IND	IND
30669	1610612747	Los Angeles Lakers	2000-06-16	LAL @ IND	IND
30670	1610612747	Los Angeles Lakers	2000-06-19	LAL vs. IND	IND
30671	1610612747	Los Angeles Lakers	2000-06-19	LAL vs. IND	IND
30672	1610612747	Los Angeles Lakers	2000-06-19	LAL vs. IND	IND
30673	1610612747	Los Angeles Lakers	2000-06-19	LAL vs. IND	IND
30674	1610612747	Los Angeles Lakers	2000-06-19	LAL vs. IND	IND
30675	1610612747	Los Angeles Lakers	2000-06-19	LAL vs. IND	IND
30676	1610612747	Los Angeles Lakers	2000-06-19	LAL vs. IND	IND
30677	1610612747	Los Angeles Lakers	2000-06-19	LAL vs. IND	IND
30678	1610612747	Los Angeles Lakers	2000-06-19	LAL vs. IND	IND
30679	1610612747	Los Angeles Lakers	2000-06-19	LAL vs. IND	IND
30681	1610612747	Los Angeles Lakers	2000-06-19	LAL vs. IND	IND
30683	1610612747	Los Angeles Lakers	2000-06-19	LAL vs. IND	IND
30684	1610612747	Los Angeles Lakers	2000-06-19	LAL vs. IND	IND
30685	1610612747	Los Angeles Lakers	2000-06-19	LAL vs. IND	IND
30687	1610612747	Los Angeles Lakers	2000-06-19	LAL vs. IND	IND
30688	1610612747	Los Angeles Lakers	2000-06-19	LAL vs. IND	IND
30689	1610612747	Los Angeles Lakers	2000-06-19	LAL vs. IND	IND
30690	1610612747	Los Angeles Lakers	2000-06-19	LAL vs. IND	IND

```
30691  1610612747  Los Angeles Lakers  2000-06-19  LAL vs. IND      IND
30692  1610612747  Los Angeles Lakers  2000-06-19  LAL vs. IND      IND
30694  1610612747  Los Angeles Lakers  2000-06-19  LAL vs. IND      IND
30695  1610612747  Los Angeles Lakers  2000-06-19  LAL vs. IND      IND
30696  1610612747  Los Angeles Lakers  2000-06-19  LAL vs. IND      IND
```

	shot_id	pred
1	2	0.0
2	3	0.0
3	4	0.0
4	5	1.0
5	6	0.0
6	7	1.0
8	9	0.0
9	10	0.0
10	11	0.0
11	12	0.0
12	13	0.0
13	14	0.0
14	15	1.0
15	16	0.0
17	18	0.0
18	19	0.0
20	21	0.0
21	22	0.0
22	23	0.0
23	24	0.0
24	25	1.0
25	26	1.0
26	27	0.0
27	28	0.0
28	29	0.0
29	30	1.0
30	31	0.0
31	32	0.0
38	39	0.0
39	40	0.0
...	...	...
30661	30662	1.0
30662	30663	1.0
30663	30664	0.0
30665	30666	1.0
30666	30667	0.0
30667	30668	0.0
30669	30670	0.0
30670	30671	1.0
30671	30672	1.0
30672	30673	0.0

```
30673   30674   0.0
30674   30675   0.0
30675   30676   0.0
30676   30677   0.0
30677   30678   0.0
30678   30679   1.0
30679   30680   1.0
30681   30682   0.0
30683   30684   0.0
30684   30685   0.0
30685   30686   0.0
30687   30688   0.0
30688   30689   0.0
30689   30690   0.0
30690   30691   0.0
30691   30692   1.0
30692   30693   1.0
30694   30695   0.0
30695   30696   0.0
30696   30697   1.0

[25697 rows x 26 columns]
```

```
In [50]: df['abs_x']= df['loc_x'].abs()
         df
```

Out[50]:

	action_type	combined_shot_type	game_event_id	game_id	\
1	Jump Shot	Jump Shot	12	20000012	
2	Jump Shot	Jump Shot	35	20000012	
3	Jump Shot	Jump Shot	43	20000012	
4	Driving Dunk Shot	Dunk	155	20000012	
5	Jump Shot	Jump Shot	244	20000012	
6	Layup Shot	Layup	251	20000012	
8	Jump Shot	Jump Shot	265	20000012	
9	Running Jump Shot	Jump Shot	294	20000012	
10	Jump Shot	Jump Shot	309	20000012	
11	Jump Shot	Jump Shot	4	20000019	
12	Running Jump Shot	Jump Shot	27	20000019	
13	Jump Shot	Jump Shot	66	20000019	
14	Jump Shot	Jump Shot	80	20000019	
15	Jump Shot	Jump Shot	86	20000019	
17	Jump Shot	Jump Shot	138	20000019	
18	Jump Shot	Jump Shot	244	20000019	
20	Jump Shot	Jump Shot	255	20000019	
21	Jump Shot	Jump Shot	265	20000019	
22	Running Jump Shot	Jump Shot	274	20000019	
23	Running Jump Shot	Jump Shot	299	20000019	
24	Running Jump Shot	Jump Shot	307	20000019	

```
25             Layup Shot             Layup            332  20000019
26              Jump Shot          Jump Shot            345  20000019
27              Jump Shot          Jump Shot            369  20000019
28              Jump Shot          Jump Shot            400  20000019
29              Jump Shot          Jump Shot            429  20000019
30      Running Jump Shot          Jump Shot            488  20000019
31              Jump Shot          Jump Shot            499  20000019
38              Jump Shot          Jump Shot            184  20000047
39              Jump Shot          Jump Shot            202  20000047
...                   ...                ...            ...       ...
30661      Slam Dunk Shot               Dunk            245  49900087
30662           Jump Shot          Jump Shot            259  49900087
30663           Jump Shot          Jump Shot            270  49900087
30665          Layup Shot             Layup            280  49900087
30666           Jump Shot          Jump Shot            295  49900087
30667           Jump Shot          Jump Shot            368  49900087
30669           Jump Shot          Jump Shot            425  49900087
30670   Running Jump Shot          Jump Shot             15  49900088
30671  Driving Layup Shot             Layup             25  49900088
30672           Jump Shot          Jump Shot             29  49900088
30673           Jump Shot          Jump Shot             36  49900088
30674           Jump Shot          Jump Shot             81  49900088
30675           Jump Shot          Jump Shot             84  49900088
30676   Running Jump Shot          Jump Shot             98  49900088
30677           Jump Shot          Jump Shot            101  49900088
30678  Driving Layup Shot             Layup            181  49900088
30679          Layup Shot             Layup            212  49900088
30681           Jump Shot          Jump Shot            218  49900088
30683           Jump Shot          Jump Shot            228  49900088
30684           Jump Shot          Jump Shot            231  49900088
30685           Jump Shot          Jump Shot            249  49900088
30687           Jump Shot          Jump Shot            284  49900088
30688           Jump Shot          Jump Shot            308  49900088
30689           Jump Shot          Jump Shot            326  49900088
30690           Jump Shot          Jump Shot            331  49900088
30691  Driving Layup Shot             Layup            382  49900088
30692           Jump Shot          Jump Shot            397  49900088
30694   Running Jump Shot          Jump Shot            426  49900088
30695           Jump Shot          Jump Shot            448  49900088
30696           Jump Shot          Jump Shot            471  49900088

          lat  loc_x  loc_y       lon  minutes_remaining  period  ...    \
1     34.0443   -157      0 -118.4268                 10       1  ...
2     33.9093   -101    135 -118.3708                  7       1  ...
3     33.8693    138    175 -118.1318                  6       1  ...
4     34.0443      0      0 -118.2698                  6       2  ...
5     34.0553   -145    -11 -118.4148                  9       3  ...
6     34.0443      0      0 -118.2698                  8       3  ...
```

8	33.9363	-65	108	-118.3348	6	3	...
9	33.9193	-33	125	-118.3028	3	3	...
10	33.8063	-94	238	-118.3638	1	3	...
11	33.9173	121	127	-118.1488	11	1	...
12	33.9343	-67	110	-118.3368	7	1	...
13	34.0403	-94	4	-118.3638	2	1	...
14	33.9973	-23	47	-118.2928	1	1	...
15	33.8523	62	192	-118.2078	0	1	...
17	33.8183	-117	226	-118.3868	8	2	...
18	33.9473	-132	97	-118.4018	11	3	...
20	33.9003	3	144	-118.2668	10	3	...
21	33.9173	134	127	-118.1358	9	3	...
22	33.9343	-16	110	-118.2858	7	3	...
23	33.8943	-109	150	-118.3788	5	3	...
24	33.9813	-46	63	-118.3158	5	3	...
25	34.0443	0	0	-118.2698	2	3	...
26	33.8483	-58	196	-118.3278	2	3	...
27	33.8583	-183	186	-118.4528	0	3	...
28	33.8713	85	173	-118.1848	8	4	...
29	33.9573	3	87	-118.2668	6	4	...
30	34.0403	121	4	-118.1488	1	4	...
31	34.0103	127	34	-118.1428	0	4	...
38	33.8603	91	184	-118.1788	3	2	...
39	33.7723	-27	272	-118.2968	0	2	...
...	...	...	...	...	...	...	...
30661	34.0443	0	0	-118.2698	9	3	...
30662	33.9913	-8	53	-118.2778	8	3	...
30663	34.0193	106	25	-118.1638	6	3	...
30665	34.0263	-14	18	-118.2838	5	3	...
30666	33.8733	-81	171	-118.3508	4	3	...
30667	33.7943	40	250	-118.2298	9	4	...
30669	33.9913	171	53	-118.0988	3	4	...
30670	34.0283	-74	16	-118.3438	9	1	...
30671	34.0443	0	0	-118.2698	8	1	...
30672	33.9893	89	55	-118.1808	8	1	...
30673	34.0443	117	0	-118.1528	7	1	...
30674	33.8283	117	216	-118.1528	2	1	...
30675	33.8283	-134	216	-118.4038	2	1	...
30676	34.0443	-141	0	-118.4108	0	1	...
30677	33.9013	-113	143	-118.3828	0	1	...
30678	34.0283	14	16	-118.2558	3	2	...
30679	34.0443	0	0	-118.2698	0	2	...
30681	33.7833	-18	261	-118.2878	0	2	...
30683	33.8283	1	216	-118.2688	10	3	...
30684	33.9553	-96	89	-118.3658	10	3	...
30685	33.7943	81	250	-118.1888	7	3	...
30687	33.9443	40	100	-118.2298	3	3	...
30688	33.9833	-126	61	-118.3958	1	3	...

```
30689  33.3653    -12    679 -118.2818                0       3  ...
30690  33.9443   -113    100 -118.3828               11       4  ...
30691  34.0443      0      0 -118.2698                7       4  ...
30692  33.9963      1     48 -118.2688                6       4  ...
30694  33.8783   -134    166 -118.4038                3       4  ...
30695  33.7773     31    267 -118.2388                2       4  ...
30696  33.9723      1     72 -118.2688                0       4  ...

             shot_zone_basic  shot_zone_range     team_id           team_name  \
1                  Mid-Range        8-16 ft.  1610612747  Los Angeles Lakers
2                  Mid-Range       16-24 ft.  1610612747  Los Angeles Lakers
3                  Mid-Range       16-24 ft.  1610612747  Los Angeles Lakers
4              Restricted Area  Less Than 8 ft.  1610612747  Los Angeles Lakers
5                  Mid-Range        8-16 ft.  1610612747  Los Angeles Lakers
6              Restricted Area  Less Than 8 ft.  1610612747  Los Angeles Lakers
8       In The Paint (Non-RA)        8-16 ft.  1610612747  Los Angeles Lakers
9       In The Paint (Non-RA)        8-16 ft.  1610612747  Los Angeles Lakers
10          Above the Break 3         24+ ft.  1610612747  Los Angeles Lakers
11                 Mid-Range       16-24 ft.  1610612747  Los Angeles Lakers
12      In The Paint (Non-RA)        8-16 ft.  1610612747  Los Angeles Lakers
13                 Mid-Range        8-16 ft.  1610612747  Los Angeles Lakers
14      In The Paint (Non-RA)  Less Than 8 ft.  1610612747  Los Angeles Lakers
15                 Mid-Range       16-24 ft.  1610612747  Los Angeles Lakers
17          Above the Break 3         24+ ft.  1610612747  Los Angeles Lakers
18                 Mid-Range       16-24 ft.  1610612747  Los Angeles Lakers
20                 Mid-Range        8-16 ft.  1610612747  Los Angeles Lakers
21                 Mid-Range       16-24 ft.  1610612747  Los Angeles Lakers
22      In The Paint (Non-RA)        8-16 ft.  1610612747  Los Angeles Lakers
23                 Mid-Range       16-24 ft.  1610612747  Los Angeles Lakers
24      In The Paint (Non-RA)  Less Than 8 ft.  1610612747  Los Angeles Lakers
25             Restricted Area  Less Than 8 ft.  1610612747  Los Angeles Lakers
26                 Mid-Range       16-24 ft.  1610612747  Los Angeles Lakers
27          Above the Break 3         24+ ft.  1610612747  Los Angeles Lakers
28                 Mid-Range       16-24 ft.  1610612747  Los Angeles Lakers
29      In The Paint (Non-RA)        8-16 ft.  1610612747  Los Angeles Lakers
30                 Mid-Range        8-16 ft.  1610612747  Los Angeles Lakers
31                 Mid-Range        8-16 ft.  1610612747  Los Angeles Lakers
38                 Mid-Range       16-24 ft.  1610612747  Los Angeles Lakers
39          Above the Break 3         24+ ft.  1610612747  Los Angeles Lakers
...                       ...             ...         ...                 ...
30661          Restricted Area  Less Than 8 ft.  1610612747  Los Angeles Lakers
30662   In The Paint (Non-RA)  Less Than 8 ft.  1610612747  Los Angeles Lakers
30663              Mid-Range        8-16 ft.  1610612747  Los Angeles Lakers
30665          Restricted Area  Less Than 8 ft.  1610612747  Los Angeles Lakers
30666              Mid-Range       16-24 ft.  1610612747  Los Angeles Lakers
30667       Above the Break 3         24+ ft.  1610612747  Los Angeles Lakers
30669              Mid-Range       16-24 ft.  1610612747  Los Angeles Lakers
30670   In The Paint (Non-RA)  Less Than 8 ft.  1610612747  Los Angeles Lakers
```

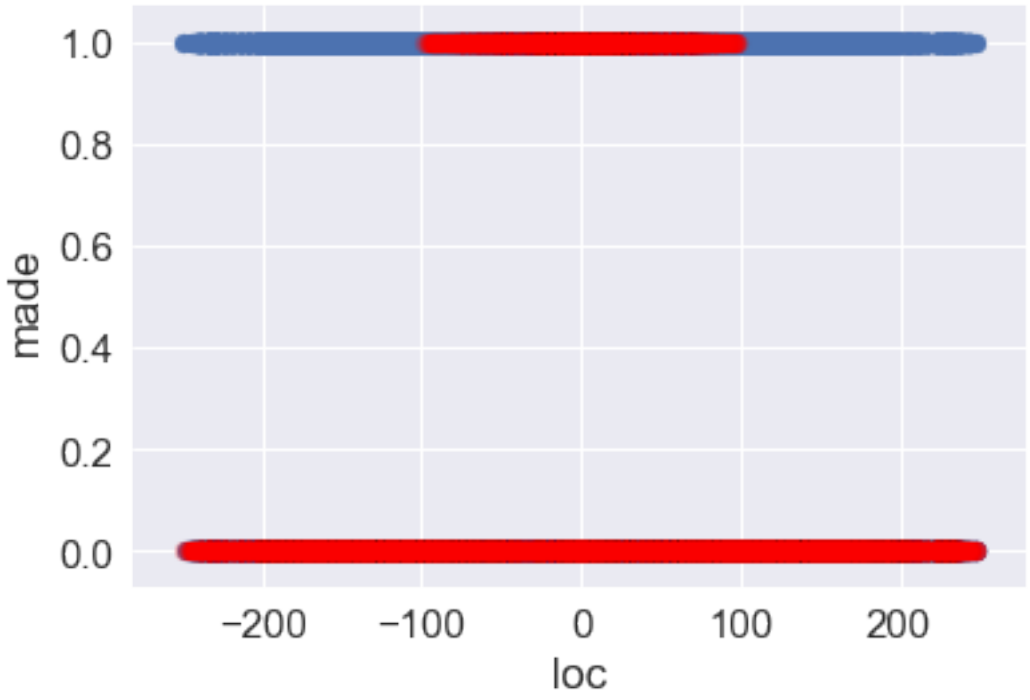
30671	Restricted Area	Less Than 8 ft.	1610612747	Los Angeles Lakers
30672	Mid-Range	8-16 ft.	1610612747	Los Angeles Lakers
30673	Mid-Range	8-16 ft.	1610612747	Los Angeles Lakers
30674	Above the Break 3	24+ ft.	1610612747	Los Angeles Lakers
30675	Above the Break 3	24+ ft.	1610612747	Los Angeles Lakers
30676	Mid-Range	8-16 ft.	1610612747	Los Angeles Lakers
30677	Mid-Range	16-24 ft.	1610612747	Los Angeles Lakers
30678	Restricted Area	Less Than 8 ft.	1610612747	Los Angeles Lakers
30679	Restricted Area	Less Than 8 ft.	1610612747	Los Angeles Lakers
30681	Above the Break 3	24+ ft.	1610612747	Los Angeles Lakers
30683	Mid-Range	16-24 ft.	1610612747	Los Angeles Lakers
30684	Mid-Range	8-16 ft.	1610612747	Los Angeles Lakers
30685	Above the Break 3	24+ ft.	1610612747	Los Angeles Lakers
30687	In The Paint (Non-RA)	8-16 ft.	1610612747	Los Angeles Lakers
30688	Mid-Range	8-16 ft.	1610612747	Los Angeles Lakers
30689	Backcourt	Back Court Shot	1610612747	Los Angeles Lakers
30690	Mid-Range	8-16 ft.	1610612747	Los Angeles Lakers
30691	Restricted Area	Less Than 8 ft.	1610612747	Los Angeles Lakers
30692	In The Paint (Non-RA)	Less Than 8 ft.	1610612747	Los Angeles Lakers
30694	Mid-Range	16-24 ft.	1610612747	Los Angeles Lakers
30695	Above the Break 3	24+ ft.	1610612747	Los Angeles Lakers
30696	In The Paint (Non-RA)	Less Than 8 ft.	1610612747	Los Angeles Lakers

	game_date	matchup	opponent	shot_id	pred	abs_x
1	2000-10-31	LAL @ POR	POR	2	0.0	157
2	2000-10-31	LAL @ POR	POR	3	0.0	101
3	2000-10-31	LAL @ POR	POR	4	0.0	138
4	2000-10-31	LAL @ POR	POR	5	1.0	0
5	2000-10-31	LAL @ POR	POR	6	0.0	145
6	2000-10-31	LAL @ POR	POR	7	1.0	0
8	2000-10-31	LAL @ POR	POR	9	0.0	65
9	2000-10-31	LAL @ POR	POR	10	0.0	33
10	2000-10-31	LAL @ POR	POR	11	0.0	94
11	2000-11-01	LAL vs. UTA	UTA	12	0.0	121
12	2000-11-01	LAL vs. UTA	UTA	13	0.0	67
13	2000-11-01	LAL vs. UTA	UTA	14	0.0	94
14	2000-11-01	LAL vs. UTA	UTA	15	1.0	23
15	2000-11-01	LAL vs. UTA	UTA	16	0.0	62
17	2000-11-01	LAL vs. UTA	UTA	18	0.0	117
18	2000-11-01	LAL vs. UTA	UTA	19	0.0	132
20	2000-11-01	LAL vs. UTA	UTA	21	0.0	3
21	2000-11-01	LAL vs. UTA	UTA	22	0.0	134
22	2000-11-01	LAL vs. UTA	UTA	23	0.0	16
23	2000-11-01	LAL vs. UTA	UTA	24	0.0	109
24	2000-11-01	LAL vs. UTA	UTA	25	1.0	46
25	2000-11-01	LAL vs. UTA	UTA	26	1.0	0
26	2000-11-01	LAL vs. UTA	UTA	27	0.0	58
27	2000-11-01	LAL vs. UTA	UTA	28	0.0	183

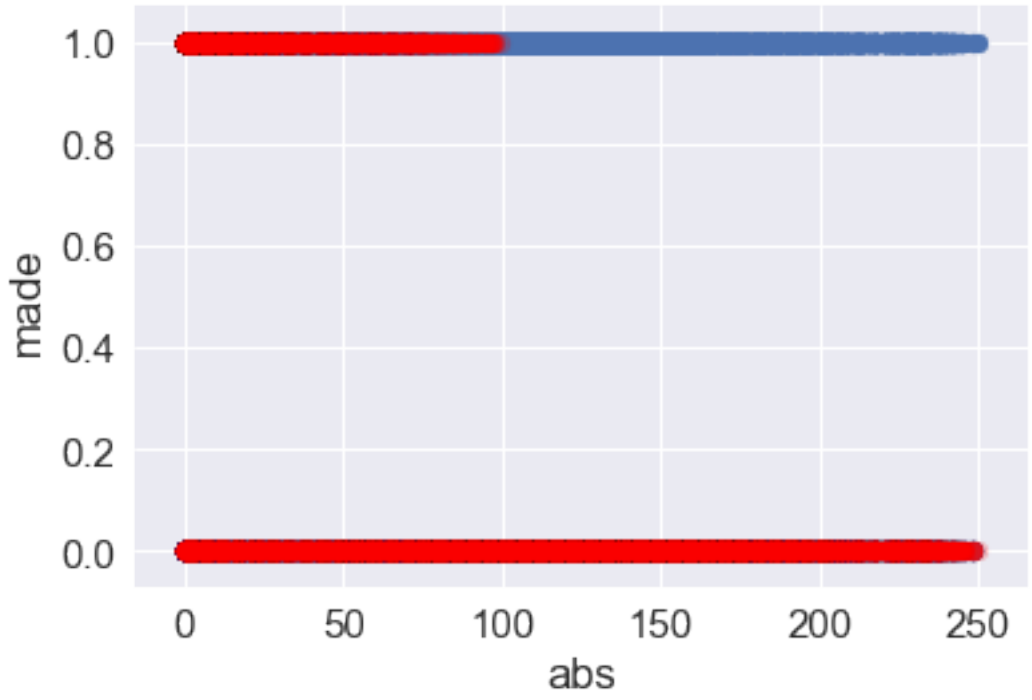
```
28     2000-11-01  LAL vs. UTA      UTA      29  0.0     85
29     2000-11-01  LAL vs. UTA      UTA      30  1.0      3
30     2000-11-01  LAL vs. UTA      UTA      31  0.0    121
31     2000-11-01  LAL vs. UTA      UTA      32  0.0    127
38     2000-11-04    LAL @ VAN      VAN      39  0.0     91
39     2000-11-04    LAL @ VAN      VAN      40  0.0     27
...           ...          ...      ...     ...  ...    ...
30661  2000-06-16    LAL @ IND      IND   30662  1.0      0
30662  2000-06-16    LAL @ IND      IND   30663  1.0      8
30663  2000-06-16    LAL @ IND      IND   30664  0.0    106
30665  2000-06-16    LAL @ IND      IND   30666  1.0     14
30666  2000-06-16    LAL @ IND      IND   30667  0.0     81
30667  2000-06-16    LAL @ IND      IND   30668  0.0     40
30669  2000-06-16    LAL @ IND      IND   30670  0.0    171
30670  2000-06-19  LAL vs. IND      IND   30671  1.0     74
30671  2000-06-19  LAL vs. IND      IND   30672  1.0      0
30672  2000-06-19  LAL vs. IND      IND   30673  0.0     89
30673  2000-06-19  LAL vs. IND      IND   30674  0.0    117
30674  2000-06-19  LAL vs. IND      IND   30675  0.0    117
30675  2000-06-19  LAL vs. IND      IND   30676  0.0    134
30676  2000-06-19  LAL vs. IND      IND   30677  0.0    141
30677  2000-06-19  LAL vs. IND      IND   30678  0.0    113
30678  2000-06-19  LAL vs. IND      IND   30679  1.0     14
30679  2000-06-19  LAL vs. IND      IND   30680  1.0      0
30681  2000-06-19  LAL vs. IND      IND   30682  0.0     18
30683  2000-06-19  LAL vs. IND      IND   30684  0.0      1
30684  2000-06-19  LAL vs. IND      IND   30685  0.0     96
30685  2000-06-19  LAL vs. IND      IND   30686  0.0     81
30687  2000-06-19  LAL vs. IND      IND   30688  0.0     40
30688  2000-06-19  LAL vs. IND      IND   30689  0.0    126
30689  2000-06-19  LAL vs. IND      IND   30690  0.0     12
30690  2000-06-19  LAL vs. IND      IND   30691  0.0    113
30691  2000-06-19  LAL vs. IND      IND   30692  1.0      0
30692  2000-06-19  LAL vs. IND      IND   30693  1.0      1
30694  2000-06-19  LAL vs. IND      IND   30695  0.0    134
30695  2000-06-19  LAL vs. IND      IND   30696  0.0     31
30696  2000-06-19  LAL vs. IND      IND   30697  1.0      1

[25697 rows x 27 columns]
```

In [53]:
```
plt.scatter(kobe.loc_x, kobe.shot_made_flag)
plt.scatter(kobe.loc_x, kobe.pred, color='red', alpha=.2)
plt.xlabel('loc')
plt.ylabel('made')
```

Out[53]:
```
Text(0,0.5,'made')
```




```
In [51]: plt.scatter(kobe.abs_x, kobe.shot_made_flag)
         plt.scatter(kobe.abs_x, kobe.pred, color='red', alpha=.2)
         plt.xlabel('abs')
         plt.ylabel('made')

Out[51]: Text(0,0.5,'made')
```



In [ ]: The absolute value shows a more accurate depiction of shots that are made and missed. T

2.0.3 2. Convert several (including ' ') string columns/features into numerical and attempt to use them in fitting a Logistic Regression model. Show histograms (similar to ones above) of made/missed of these new numerical features. Use these histograms to explain and justify why these features could improve the model

In [54]:
```
df.dtypes
```

Out[54]:
```
action_type            object
combined_shot_type     object
game_event_id           int64
game_id                 int64
lat                   float64
loc_x                   int64
loc_y                   int64
lon                   float64
minutes_remaining       int64
period                  int64
playoffs                int64
season                 object
seconds_remaining       int64
shot_distance           int64
shot_made_flag        float64
```

```
shot_type              object
shot_zone_area         object
shot_zone_basic        object
shot_zone_range        object
team_id                 int64
team_name              object
game_date              object
matchup                object
opponent               object
shot_id                 int64
pred                  float64
abs_x                   int64
dtype: object
```

In [61]:
```
df["shot_zone_area"] = pd.to_numeric(df["shot_zone_area"])
df["shot_zone_basic"] = pd.to_numeric(df["shot_zone_basic"])
df["shot_zone_range"] = pd.to_numeric(df["shot_zone_range"])

df[["shot_zone_area", "shot_zone_basic", "shot_zone_range"]] = df[["shot_zone_area",
```

```
---------------------------------------------------------------------------

ValueError                                Traceback (most recent call last)

pandas/_libs/src\inference.pyx in pandas._libs.lib.maybe_convert_numeric()


ValueError: Unable to parse string "Left Side(L)"


During handling of the above exception, another exception occurred:


ValueError                                Traceback (most recent call last)

<ipython-input-61-553dc397a3c6> in <module>()
----> 1 df["shot_zone_area"] = pd.to_numeric(df["shot_zone_area"])
      2 df["shot_zone_basic"] = pd.to_numeric(df["shot_zone_basic"])
      3 df["shot_zone_range"] = pd.to_numeric(df["shot_zone_range"])
      4
      5 df[["shot_zone_area", "shot_zone_basic", "shot_zone_range"]] = df[["shot_zone_area"


~\Anaconda3\lib\site-packages\pandas\core\tools\numeric.py in to_numeric(arg, errors, c
131             coerce_numeric = False if errors in ('ignore', 'raise') else True
132             values = lib.maybe_convert_numeric(values, set(),
--> 133                                                coerce_numeric=coerce_numeric)
```

```
134
135     except Exception:


pandas/_libs/src\inference.pyx in pandas._libs.lib.maybe_convert_numeric()


ValueError: Unable to parse string "Left Side(L)" at position 0
```

```
In [71]: from sklearn.linear_model import LogisticRegression as Model
         # from sklearn.tree import DecisionTreeClassifier as Model
         # from sklearn.ensemble import RandomForestClassifier as Model
         model = Model()

         from sklearn.metrics import (accuracy_score,
                                      classification_report,
                                      confusion_matrix, auc, roc_curve
                                     )
         from sklearn.metrics import *
         from sklearn import cross_validation

         X_train, X_test, y_train, y_test = cross_validation.train_test_split(
             X, y, test_size=0.4, random_state=0)

         cross_validation.cross_val_score(model, X, y, cv=10)

         model.fit(X,y)
         kobe['pred']=model.predict(X)
```

2.0.4 3. Show a 3 dimensional surface plot [https://matplotlib.org/mpl_toolkits/mplot3d/tutorial.html#surface plots] of probabilities from a trained Logistic Regression model using only `abs_x` and `loc_y`. The probabilities arise from a distributed grid of `x` values and `y` values as input to the `predict_proba()` function.

```
In [73]: fig_x=kobe['abs_x']
         fig_y=kobe['loc_y']
         x_train = np.column_stack ((model.predict(X), fig_x))
         y_train = np.column_stack ((model.predict(X), fig_y))


         import matplotlib.pyplot as plt
         from mpl_toolkits.mplot3d import Axes3D
         fig = plt.figure()
         ax = fig.add_subplot(111, projection='3d')


         fig_x= x_train
```

```
fig_y= y_train
Z=model.predict_proba(x_train)
```