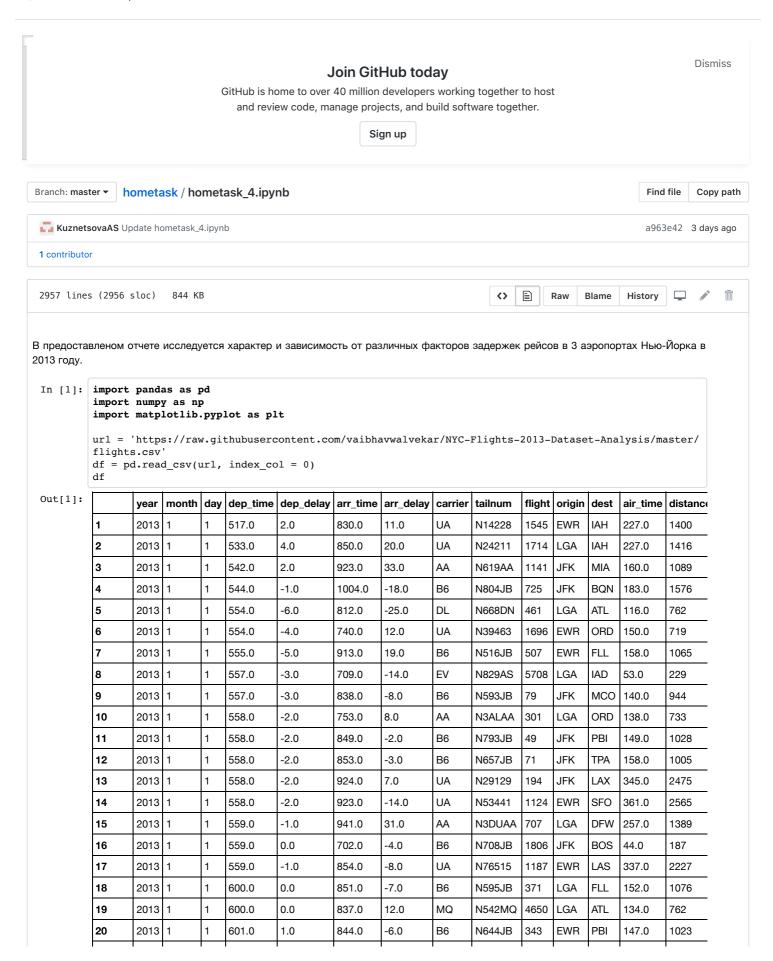
KuznetsovaAS / hometask



21	2013	1	1	602.0	-8.0	812.0	-8.0	DL	N971DL	1919	LGA	MSP	170.0	1020
22	2013	1	1	602.0	-3.0	821.0	16.0	MQ	N730MQ	4401	LGA	DTW	105.0	502
23	2013	1	1	606.0	-4.0	858.0	-12.0	AA	N633AA	1895	EWR	MIA	152.0	1085
24	2013	1	1	606.0	-4.0	837.0	-8.0	DL	N3739P	1743	JFK	ATL	128.0	760
25	2013	1	1	607.0	0.0	858.0	-17.0	UA	N53442	1077	EWR	MIA	157.0	1085
26	2013	1	1	608.0	8.0	807.0	32.0	MQ	N9EAMQ	3768	EWR	ORD	139.0	719
27	2013	1	1	611.0	11.0	945.0	14.0	UA	N532UA	303	JFK	SFO	366.0	2586
28	2013	1	1	613.0	3.0	925.0	4.0	B6	N635JB	135	JFK	RSW	175.0	1074
29	2013	1	1	615.0	0.0	1039.0	-21.0	B6	N794JB	709	JFK	SJU	182.0	1598
30	2013	1	1	615.0	0.0	833.0	-9.0	DL	N326NB	575	EWR	ATL	120.0	746
336747	2013	9	30	2123.0	-2.0	2223.0	-24.0	EV	N712EV	5489	LGA	CHO	45.0	305
336748	2013	9	30	2127.0	-2.0	2314.0	-9.0	EV	N16546	3833	EWR	CLT	72.0	529
336749	2013	9	30	2128.0	-2.0	2328.0	-31.0	B6	N807JB	97	JFK	DEN	213.0	1626
336750	2013	9	30	2129.0	30.0	2230.0	-2.0	EV	N751EV	5048	LGA	RIC	45.0	292
336751	2013	9	30	2131.0	-9.0	2225.0	-30.0	MQ	N807MQ	3621	JFK	DCA	36.0	213
336752	2013	9	30	2140.0	0.0	10.0	-30.0	AA	N335AA	185	JFK	LAX	298.0	2475
336753	2013	9	30	2142.0	13.0	2250.0	11.0	EV	N12957	4509	EWR	PWM	47.0	284
336754	2013	9	30	2145.0	0.0	115.0	-25.0	B6	N633JB	1103	JFK	SJU	192.0	1598
336755	2013	9	30	2147.0	10.0	30.0	3.0	B6	N627JB	1371	LGA	FLL	139.0	1076
336756	2013	9	30	2149.0	-7.0	2245.0	-23.0	UA	N813UA	523	EWR	BOS	37.0	200
336757	2013	9	30	2150.0	-9.0	2250.0	-16.0	EV	N10575	3842	EWR	MHT	39.0	209
336758	2013	9	30	2159.0	194.0	2344.0	194.0	9E	N906XJ	3320	JFK	BUF	50.0	301
336759	2013	9	30	2203.0	-2.0	2339.0	8.0	EV	N722EV	5311	LGA	BGR	61.0	378
336760	2013	9	30	2207.0	27.0	2257.0	7.0	MQ	N532MQ	3660	LGA	BNA	97.0	764
336761	2013	9	30	2211.0	72.0	2339.0	57.0	EV	N12145	4672	EWR	STL	120.0	872
336762	2013	9	30	2231.0	-14.0	2335.0	-21.0	B6	N193JB	108	JFK	PWM	48.0	273
336763	2013	9	30	2233.0	80.0	112.0	42.0	UA	N578UA	471	EWR	SFO	318.0	2565
336764	2013	9	30	2235.0	154.0	59.0	130.0	B6	N804JB	1083	JFK	МСО	123.0	944
336765	2013	9	30	2237.0	-8.0	2345.0	-8.0	B6	N318JB	234	JFK	BTV	43.0	266
336766	2013	9	30	2240.0	-5.0	2334.0	-17.0	B6	N354JB	1816	JFK	SYR	41.0	209
336767	2013	9	30	2240.0	-10.0	2347.0	-20.0	B6	N281JB	2002	JFK	BUF	52.0	301
336768	2013	9	30	2241.0	-5.0	2345.0	-16.0	B6	N346JB	486	JFK	ROC	47.0	264
336769	2013	9	30	2307.0	12.0	2359.0	1.0	B6	N565JB	718	JFK	BOS	33.0	187
336770	2013	9	30	2349.0	-10.0	325.0	-25.0	B6	N516JB	745	JFK	PSE	196.0	1617
336771	2013	9	30	NaN	NaN	NaN	NaN	EV	N740EV	5274	LGA	BNA	NaN	764
336772	2013	9	30	NaN	NaN	NaN	NaN	9E	NaN	3393	JFK	DCA	NaN	213
336773	2013	9	30	NaN	NaN	NaN	NaN	9E	NaN	3525	LGA	SYR	NaN	198
336774	2013	9	30	NaN	NaN	NaN	NaN	MQ	N535MQ	3461	LGA	BNA	NaN	764
336775	2013	9	30	NaN	NaN	NaN	NaN	MQ	N511MQ	3572	LGA	CLE	NaN	419
336776	2013	9	30	NaN	NaN	NaN	NaN	MQ	N839MQ	3531	LGA	RDU	NaN	431

336776 rows × 16 columns

```
In [2]:     url1 = 'weather.txt'
     w = pd.read_csv(url1, index_col = 0)
     w
```

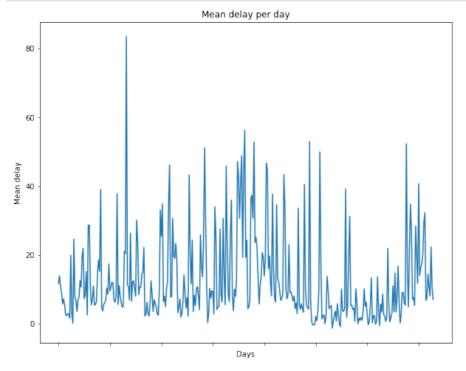
Out[2]: origin year month day hour temp dewp humid wind_dir wind_speed wind_gust precip pressure visib

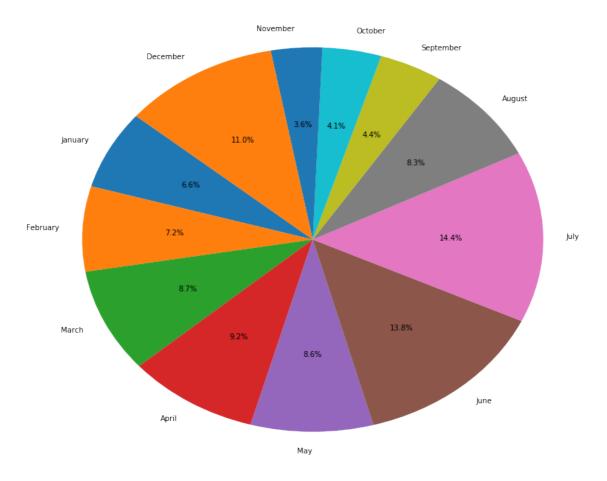
1	EWR	2013	1	l ₁	О	37.04	21.92	53.97	230.0	10.35702	11.918651	0.0	1013.9	10.0
2	EWR	2013		1	1	37.04			230.0	13.80936	15.891535	0.0	1013.0	10.0
3	EWR	2013		1	2	37.94	21.92		230.0	12.65858	14.567241	0.0	1012.6	10.0
4	EWR	2013		1	3	37.94	23.00	54.51	230.0	13.80936	15.891535	0.0	1012.7	10.0
5	EWR	2013	1	1	4	37.94	24.08	57.04	240.0	14.96014	17.215830	0.0	1012.8	10.0
6	EWR	2013		1	6	39.02	26.06	59.37	270.0	10.35702	11.918651	0.0	1012.0	10.0
7	EWR	2013		1	7	39.02	26.96		250.0	8.05546	9.270062	0.0	1012.3	10.0
8	EWR	2013	1	1	8	39.02	28.04	64.43	240.0	11.50780	13.242946	0.0	1012.5	10.0
9	EWR	2013		1	9	39.92	28.04	62.21	250.0	12.65858	14.567241	0.0	1012.2	10.0
10	EWR	2013	1	1	10	39.02	28.04	64.43	260.0	12.65858	14.567241	0.0	1011.9	10.0
11	EWR	2013	1	1	11	37.94	28.04	67.21	240.0	11.50780	13.242946	0.0	1012.4	10.0
12	EWR	2013	1	1	12	39.02	28.04	64.43	240.0	14.96014	17.215830	0.0	1012.2	10.0
13	EWR	2013	1	1	13	39.92	28.04	62.21	250.0	10.35702	11.918651	0.0	1012.2	10.0
14	EWR	2013	1	1	14	39.92	28.04	62.21	260.0	14.96014	17.215830	0.0	1012.7	10.0
15	EWR	2013	1	1	15	41.00	28.04	59.65	260.0	13.80936	15.891535	0.0	1012.4	10.0
16	EWR	2013	1	1	16	41.00	26.96	57.06	260.0	14.96014	17.215830	0.0	1011.4	10.0
17	EWR	2013	1	1	17	39.20	28.40	64.93	270.0	16.11092	18.540125	0.0	NaN	10.0
18	EWR	2013	1	1	18	39.20	28.40	64.93	330.0	14.96014	17.215830	0.0	NaN	10.0
19	EWR	2013	1	1	19	39.02	24.08	54.68	280.0	13.80936	15.891535	0.0	1010.8	10.0
20	EWR	2013	1	1	20	37.94	24.08	57.04	290.0	9.20624	10.594357	0.0	1011.9	10.0
21	EWR	2013	1	1	21	37.04	19.94	49.62	300.0	13.80936	15.891535	0.0	1012.1	10.0
22	EWR	2013	1	1	22	35.96	19.04	49.83	330.0	11.50780	13.242946	0.0	1013.2	10.0
23	EWR	2013	1	1	23	33.98	15.08	45.43	310.0	12.65858	14.567241	0.0	1014.1	10.0
24	EWR	2013	1	2	0	33.08	12.92	42.84	320.0	10.35702	11.918651	0.0	1014.4	10.0
25	EWR	2013	1	2	1	32.00	15.08	49.19	310.0	14.96014	17.215830	0.0	1015.2	10.0
26	EWR	2013	1	2	2	30.02	12.92	48.48	320.0	18.41248	21.188714	0.0	1016.0	10.0
27	EWR	2013	1	2	3	28.94	12.02	48.69	320.0	18.41248	21.188714	0.0	1016.5	10.0
28	EWR	2013	1	2	4	28.04	10.94	48.15	310.0	16.11092	18.540125	0.0	1016.4	10.0
29	EWR	2013	1	2	5	26.96	10.94	50.34	310.0	14.96014	17.215830	0.0	1016.3	10.0
30	EWR	2013	1	2	6	26.06	10.94	52.25	330.0	12.65858	14.567241	0.0	1016.3	10.0
8690	EWR	2013	12	30	4	42.08	35.96	78.74	280.0	13.80936	15.891535	0.0	1005.6	10.0
8691	EWR	2013	12	30	5	41.00	35.96	82.09	260.0	5.75390	6.621473	0.0	1006.5	10.0
8692	EWR	2013	12	30	6	37.94	33.98	85.49	220.0	6.90468	7.945768	0.0	1008.1	10.0
8693	EWR	2013	12	30	7	37.04	33.98	88.57	220.0	8.05546	9.270062	0.0	1008.6	10.0
8694	EWR	2013	12	30	8	37.04	35.96	95.82	230.0	4.60312	5.297178	0.0	1008.5	10.0
8695	EWR	2013	12	30	9	37.94	35.96	92.49	230.0	8.05546	9.270062	0.0	1009.5	10.0
8696	EWR	2013	12	30	10	37.04	35.06	92.46	220.0	8.05546	9.270062	0.0	1010.1	10.0
8697	EWR	2013	12	30	11	39.02	35.96	88.66	230.0	8.05546	9.270062	0.0	1011.3	10.0
8698	EWR	2013	12	30	12	39.92	37.04	89.34	240.0	8.05546	9.270062	0.0	1012.3	10.0
8699	EWR	2013	12	30	13	42.08	37.04	82.17	240.0	10.35702	11.918651	0.0	1013.2	10.0
8700	EWR	2013	12	30	14	44.96	37.04	73.59	260.0	10.35702	11.918651	0.0	1014.0	10.0
8701	EWR	2013	12	30	15	42.98	32.00	64.93	320.0	13.80936	15.891535	0.0	1015.0	10.0
8702	EWR	2013	12	30	16	41.00	28.94	61.89	330.0	14.96014	17.215830	0.0	1015.1	10.0
8703	EWR	2013	12	30	17	39.92	26.06	57.33	320.0	13.80936	15.891535	0.0	1015.9	10.0
8704	EWR	2013	12	30	18	37.94	23.00	54.51	330.0	19.56326	22.513008	0.0	1016.7	10.0
•	•	•	•	•	•		•	•	•	•			•	

8705	EWR	2013	12	30	19	37.04	21.02	51.95	320.0	17.26170	19.864419	0.0	1017.6	10.0
8706	EWR	2013	12	30	20	35.06	17.96	49.30	340.0	17.26170	19.864419	0.0	1019.1	10.0
8707	EWR	2013	12	30	21	33.08	15.98	48.98	320.0	14.96014	17.215830	0.0	1019.8	10.0
8708	EWR	2013	12	30	22	30.92	12.92	46.74	340.0	16.11092	18.540125	0.0	1020.5	10.0
8709	EWR	2013	12	30	23	28.94	12.02	48.69	330.0	14.96014	17.215830	0.0	1021.1	10.0
8710	JFK	2013	2	18	4	17.96	-0.94	42.69	290.0	29.92028	34.431660	0.0	1016.2	10.0
8711	JFK	2013	2	20	19	32.00	8.06	36.03	280.0	26.46794	30.458776	0.0	1011.2	10.0
8712	JFK	2013	7	2	11	71.60	69.80	94.06	180.0	11.50780	13.242946	0.0	NaN	0.5
8713	JFK	2013	7	2	13	71.60	69.80	94.06	190.0	10.35702	11.918651	0.0	NaN	1.0
8714	JFK	2013	7	31	6	71.06	55.04	56.93	320.0	8.05546	9.270062	0.0	1020.4	10.0
8715	JFK	2013	9	2	20	75.20	73.40	94.14	200.0	4.60312	5.297178	0.0	NaN	4.0
8716	JFK	2013	10	23	10	48.92	39.02	68.51	60.0	4.60312	5.297178	0.0	1008.1	10.0
8717	JFK	2013	10	23	11	48.92	39.02	68.51	40.0	4.60312	5.297178	0.0	1008.5	10.0
8718	JFK	2013	12	17	5	26.96	10.94	50.34	40.0	4.60312	5.297178	0.0	1023.9	10.0
8719	LGA	2013	8	22	22	75.92	66.92	73.68	210.0	8.05546	9.270062	0.0	1011.9	10.0

8719 rows × 14 columns

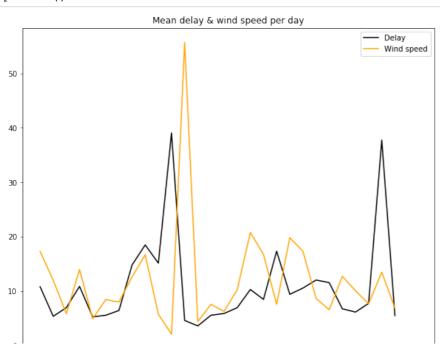
```
In [17]: # Ex1 - Средняя задержка по дням за год
plt.figure(figsize=(10,8))
    df.groupby(['month', 'day'])['dep_delay'].mean().plot()
    #plt.scatter(np.arange(365), df_mean)
    plt.title('Mean delay per day')
    plt.xlabel('Days')
    plt.ylabel('Mean delay')
plt.show()
```





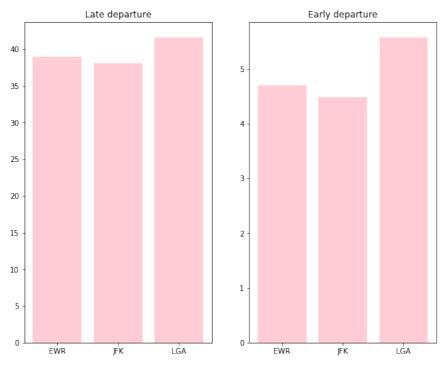
Mean delay per month

```
In [18]: # Ex3 - сравнение средней задержки и средней скорости ветра в феврале(взят для наглядности; в оста льные месяцы тенденция та же)
plt.figure(figsize=(10,8))
df[df['month']==2].groupby(['month', 'day'])['dep_delay'].mean().plot(color = 'black', label = 'De lay')
w[w['month']==2].groupby(['month', 'day'])['wind_speed'].mean().plot(color = 'orange', label = 'Wi nd speed')
plt.xticks(range(31))
plt.legend()
plt.title('Mean delay & wind speed per day')
plt.xlabel('Days')
plt.show()
```



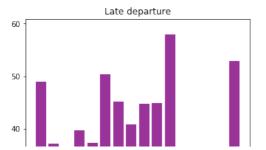
In [6]: # Ex4 - среднее значение поздних и ранних вылетов по трем аэропортам
plt.figure(figsize=(10,8))
a = df[df.dep_delay > 0].groupby('origin')['dep_delay'].mean()
b = df[df.dep_delay < 0].groupby('origin')['dep_delay'].mean()
print('bepature according to origin')
plt.subplot(1,2,1)
plt.bar(np.arange(3), a, alpha=0.8, color = 'pink')
plt.xticks(np.arange(3), ('EWR', 'JFK', 'LGA'))
plt.title('Late departure')
plt.subplot(1,2,2)
plt.bar(np.arange(3), b.abs(), alpha=0.8, color = 'pink')
plt.xticks(np.arange(3), ('EWR', 'JFK', 'LGA'))
plt.title('Early departure')
plt.show()</pre>

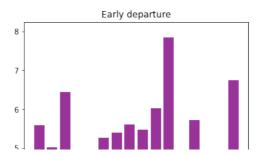
Depature according to origin

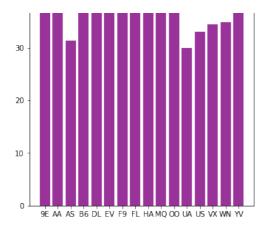


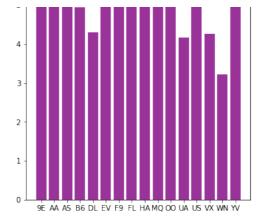
In [7]: # Ex5 – среднее значение поздних и ранних вылетов по перевозчикам plt.figure(figsize=(12,8)) a = df[df.dep_delay > 0].groupby('carrier')['dep_delay'].mean() b = df[df.dep_delay < 0].groupby('carrier')['dep_delay'].mean()</pre> print('Depature according to carrier') plt.subplot(1,2,1) plt.bar(np.arange(len(df['carrier'].unique())), a, alpha=0.8, color = 'purple') plt.xticks(np.arange(len(df['carrier'].unique())), ('9E','AA','AS','B6','DL','EV','F9','FL','HA',' MQ','OO','UA','US','VX','WN','YV')) plt.title('Late departure') plt.subplot(1,2,2)plt.bar(np.arange(len(df['carrier'].unique())), b.abs(), alpha=0.8, color = 'purple')
plt.xticks(np.arange(len(df['carrier'].unique())), ('9E','AA','AS','B6','DL','EV','F9','FL','HA',' MQ','OO','UA','US','VX','WN','YV')) plt.title('Early departure') plt.show()

Depature according to carrier

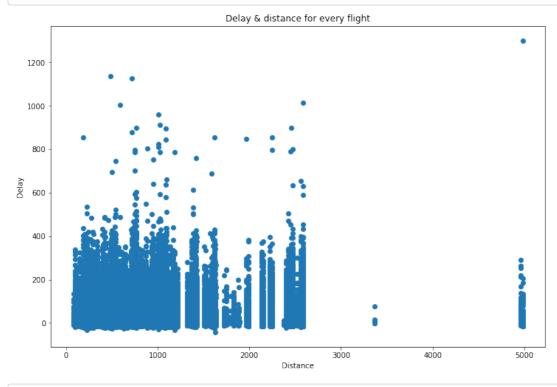




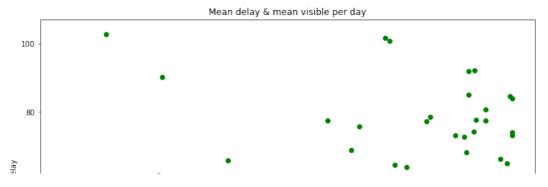


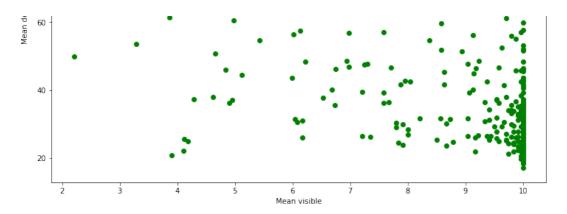


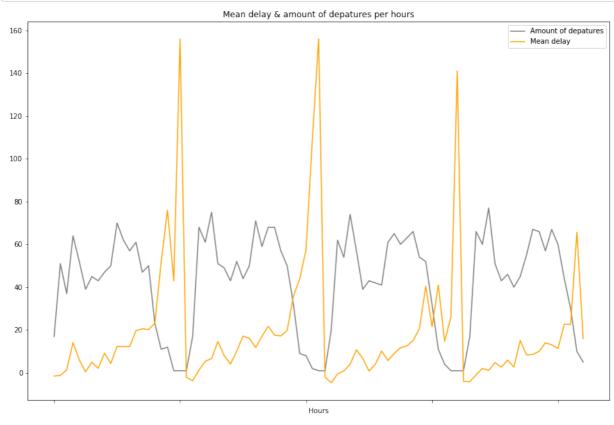
```
In [19]: # Ex6 - задержка вылета не зависит от расстояния полета
   plt.figure(figsize=(12,8))
   plt.scatter(df['distance'], df['dep_delay'])
   plt.title('Delay & distance for every flight')
   plt.xlabel('Distance')
   plt.ylabel('Delay')
   plt.show()
```



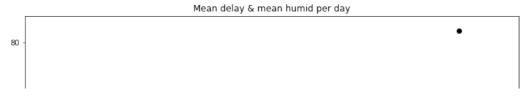
In [20]: # Ex7 - асимптотика задержки вылета при видимости 10.0
 plt.figure(figsize=(12,8))
 plt.scatter(w.groupby(['month','day'])['visib'].mean(), df[df.dep_delay > 0].groupby(['month','day '])['dep_delay'].mean()[0:364], color = 'green')
 plt.title('Mean delay & mean visible per day')
 plt.xlabel('Mean visible')
 plt.ylabel('Mean delay')
 plt.show()

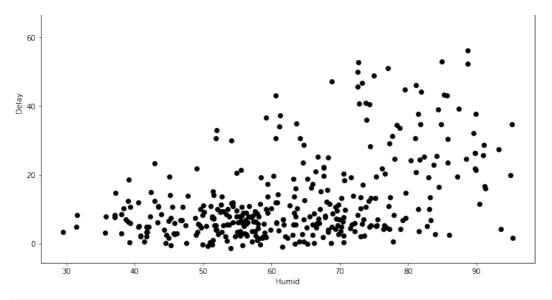




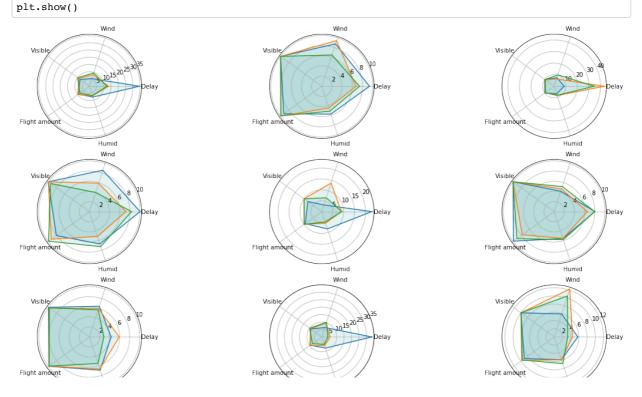


```
In [23]: # Ex9 - Зависимость задержек рейсов от влажности
plt.figure(figsize=(12,8))
plt.scatter(w.groupby(['month','day'])['humid'].mean(), df.groupby(['month','day'])['dep_delay'].m
ean()[0:364], color = 'black')
plt.title('Mean delay & mean humid per day')
plt.xlabel('Humid')
plt.ylabel('Delay')
plt.show()
```





In [24]: # Ex10 - наглядная демонстрация зависимости задержки рейсов от скорости ветра, видимости, количест ва самолетов в очереди на # вылет, влажности. Для компактности на каждом из графиков изображено по 3 дня import math dep_mean = df.groupby(['month', 'day'])['dep_delay'].mean() wind = w.groupby(['month', 'day'])['wind_speed'].mean()
vis = w.groupby(['month', 'day'])['visib'].mean() dep_c = df.groupby(['month','day'])['dep_delay'].count() h = w.groupby(['month','day'])['humid'].mean() angles = [n / float(5) * 2 * math.pi for n in range(5)]angles += angles [:1] plt.figure(figsize=(20,15)) i=8 for k in range(1,11): plt.subplot(4,3,k, polar=True) plt.xticks(angles[:-1], ['Delay','Wind','Visible','Flight amount','Humid']) for m in range(1,4): j=(k-1)*3+mvalues = [dep_mean[i][j], wind[i][j], vis[i][j], dep_c[i][j]/100, h[i][j]/10, dep_mean[i][j]] plt.plot(angles, values) plt.fill(angles, values, 'teal', alpha=0.1)





После проведенного анализа полученных данных о количестве и характере задержек рейсов, а также данных о погоде за 2013 год, можно сделать следующие выводы:

1)Средняя задержка в день в течение года постоянно колеблется в промежутке между 0 и 80 минут в день. Но средняя задержка в день преимущественно всегда не превосходит 60 минут в день.

2)Если рассматривать не среднее значение задержек рейсов по дням, а отдельно время задержки в каждом рейсе, можно заметить, что, в основном, оно находится в промежутке от 0 до 400 минут. А также можно заметить, что оно не зависит от расстояния полета.

3)Наибольшее значение средней задержки в день отмечается в летние месяцы (июнь, июль) и в декабре.

4)Сравнение средней задержки и средней скорости ветра (по дням) показало, что в среднем они имеют схожие графики подьема и спада. То есть, среднее время задержки значительно зависит от скорости ветра: чем сильнее порывы ветра, тем больше время задержки рейсов.

5)Исследование задержек рейсов по трем различным аэропортам Нью-Йорка позволяет сделать вывод, что наиболее "непунктуальны" в плане вылетов самолетов в аэропорту LGA

6)А исследование задержек рейсов по различным перевозчикам позволяет сделать вывод о том, что наиболее "непуктуальные" перевозчики - ОО и YV

7)При видимости, стремящейся к 10.0, задержка вылетов самолетов стремится к 0, в противном случае задержки не имеют асимптотики.

8)Количество самолетов, вылетающих в конкретный час, прямо пропорционально средней задержке вылетов в данный час. То есть, чем больше количество вылетов, тем больше средняя задержка в конкретный час, но она колеблется незначительно вместе с колебанием количества вылетов самолетов. Постепенно задержка увеличивается все больше и больше, и к концу дня она достигает своего максимума за этот день. Далее ночью задержка снова опускается практически до 0, поскольку ночью количество вылетов на порядок ниже.

9)Наблюдается зависимость задержек рейсов от влажности: если влажность низкая, то задержки сравнительно небольшие, а при увеличении влажности диапазон колебаний средних задержек увеличивается линейно.

10)Суммируя все полученные данные, мы наглядно продемонстрировали на последнем графике зависимость задержки рейсов от таких факторов как: скорость ветра, видимость, количество самолетов в очереди на вылет, влажность. Безусловно, задержка рейсов также зависит от множества других факторов. Но если мы рассматриваем задержку вылета в связи с метеоусловиями, и у нас в распоряжении есть ограниченное количество данных, в целом, мы можем пронаблюдать следующую картину. При увеличении скорости и порывов ветра, количества вылетающих самолетов, влажности, а также при уменьшении видимости задержка практически всегда увеличивается. Верно и обратное.