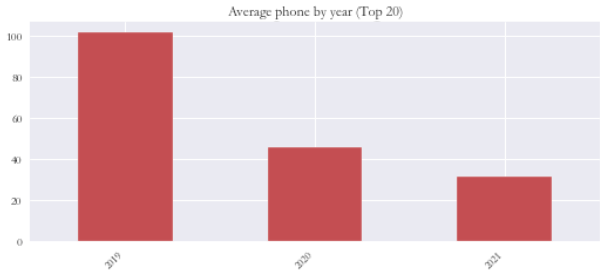
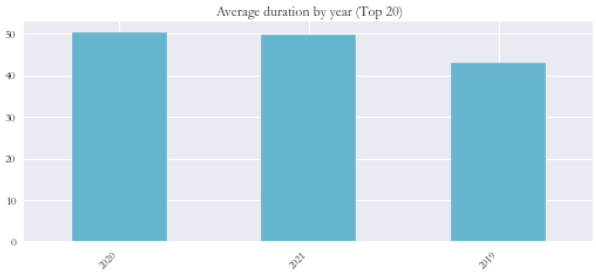
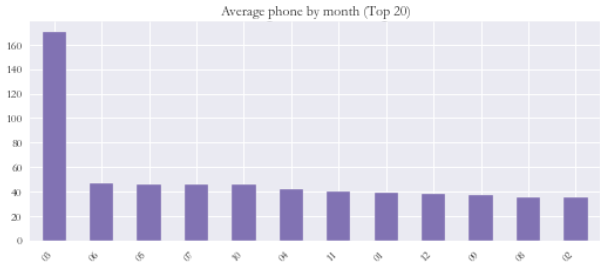
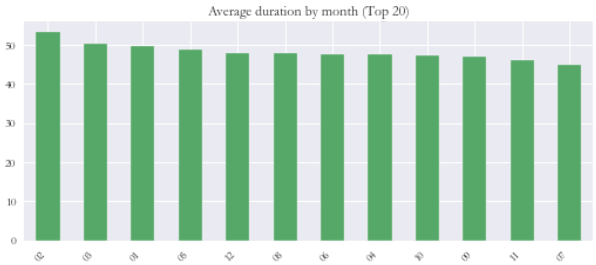
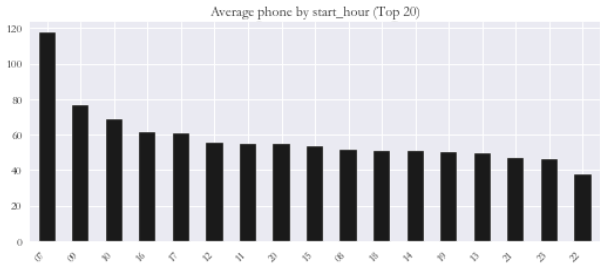
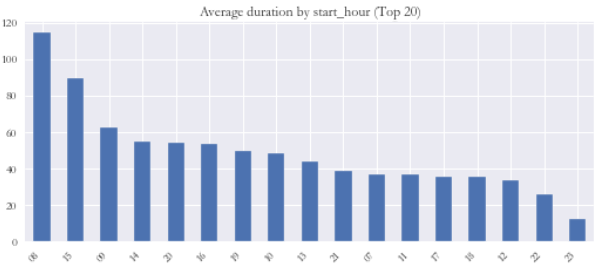

Average phone by year (Top 20)

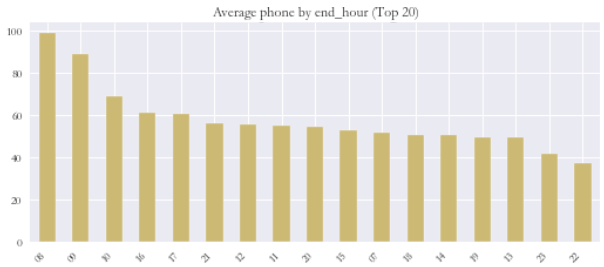

Average duration by year (Top 20)


Average phone by month (Top 20)

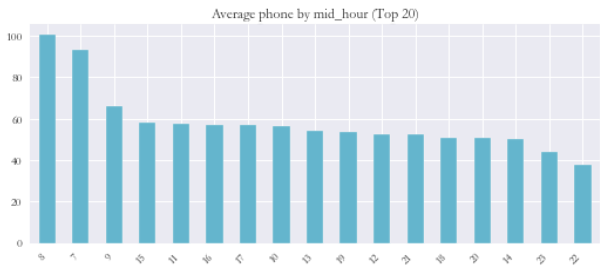

Average duration by month (Top 20)

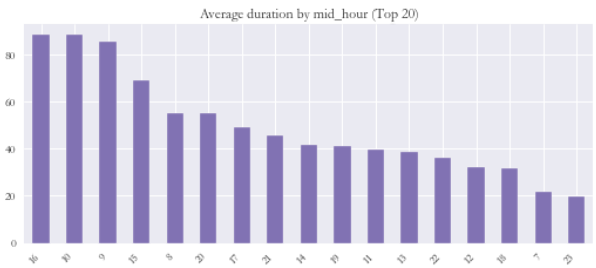

Average phone by start_hour (Top 20)

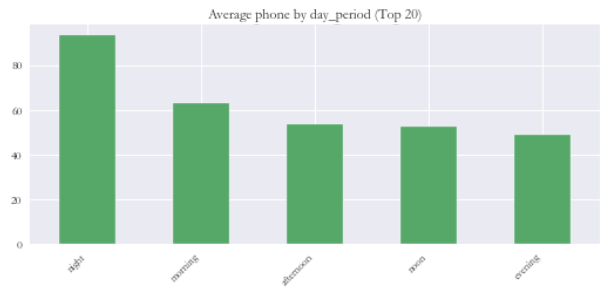

Average duration by start_hour (Top 20)

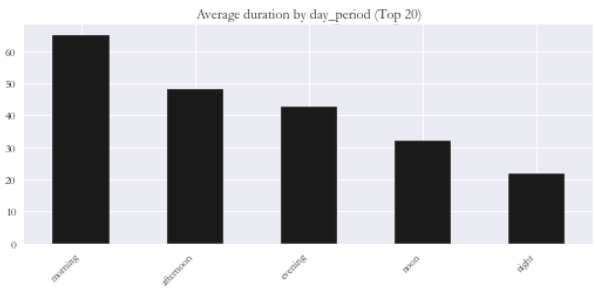

Average phone by end_hour (Top 20)

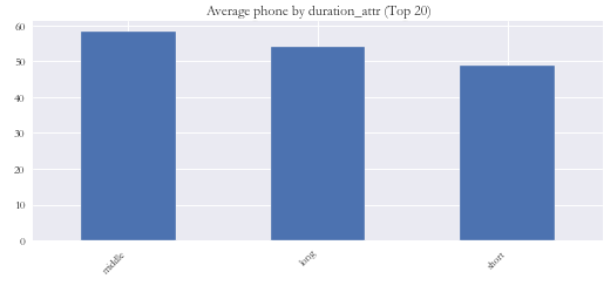

Average duration by end_hour (Top 20)

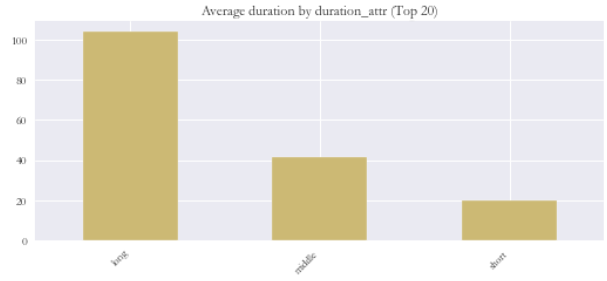

Average phone by mid_hour (Top 20)

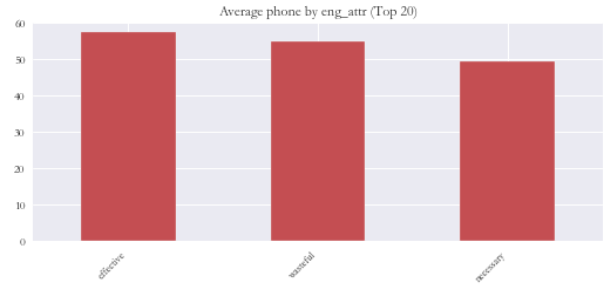

Average duration by mid_hour (Top 20)

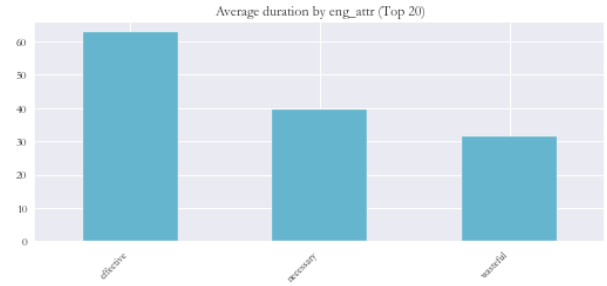

Average phone by day_period (Top 20)

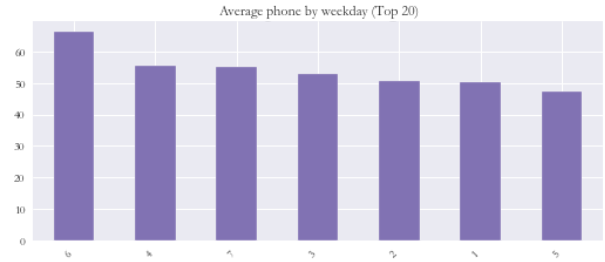

Average duration by day_period (Top 20)

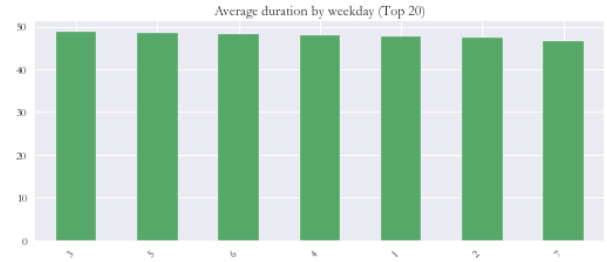


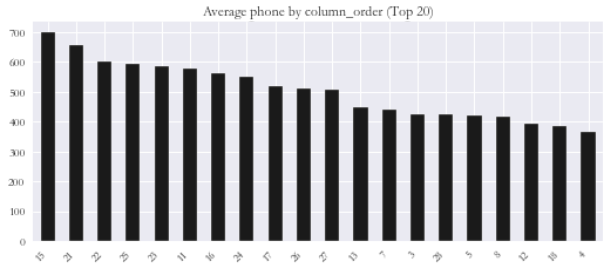


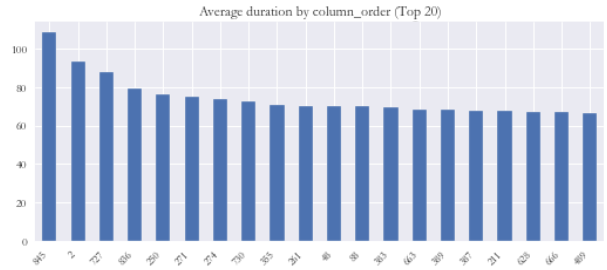













```
All Plots done
Time to run AutoViz (in seconds) = 16.150

 ###################### VISUALIZATION Completed ########################
```

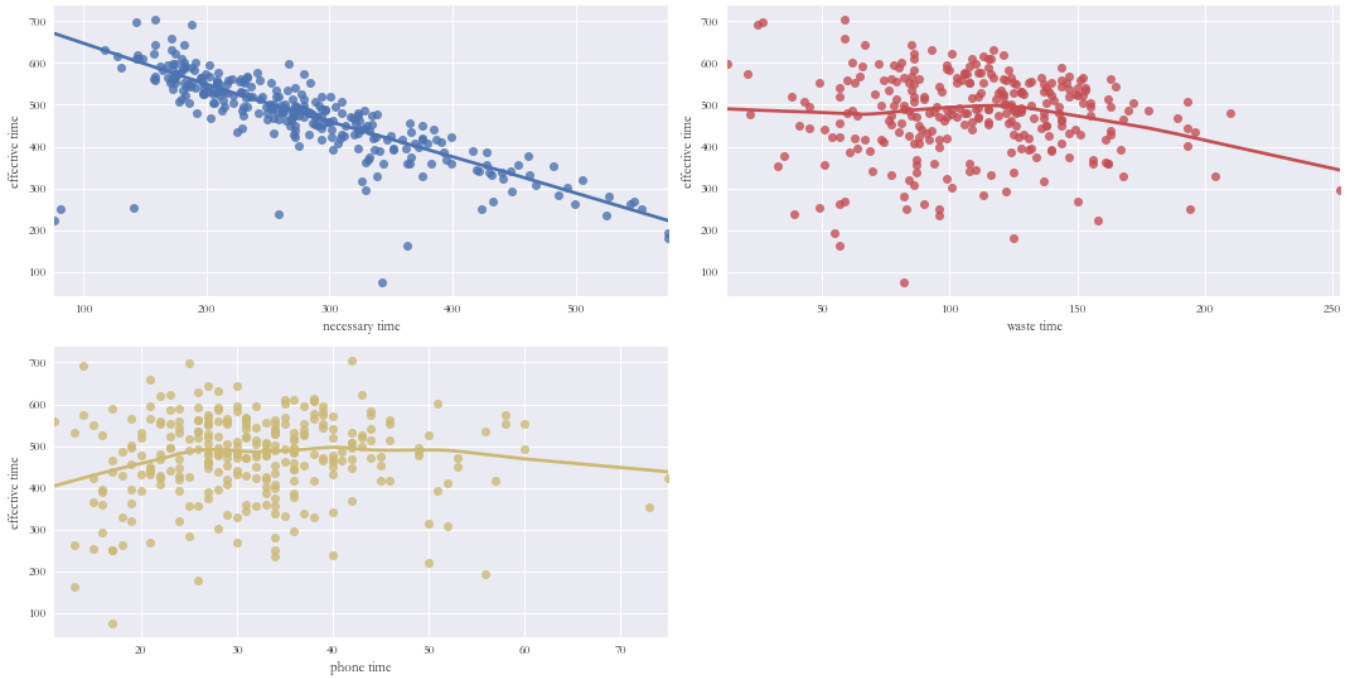
2.5.2 衍生数据

```
AV = AutoViz_Class()
dft = AV.AutoViz(
    filename=None,
    sep=",",
    depVar="effective time",
    dfte=daily_attr_copy,
    header=0,
    verbose=0,
    lowess=True,
    chart_format="svg",
    max_rows_analyzed=150000,
    max_cols_analyzed=30,
)
```

```
Shape of your Data Set loaded: (292, 9)
############## C L A S S I F Y I N G  V A R I A B L E S  ###################
Classifying variables in data set...
week_order of type=period[W-SUN] is not classified
    Number of Numeric Columns =  3
    Number of Integer-Categorical Columns =  1
    Number of String-Categorical Columns =  1
    Number of Factor-Categorical Columns =  0
```

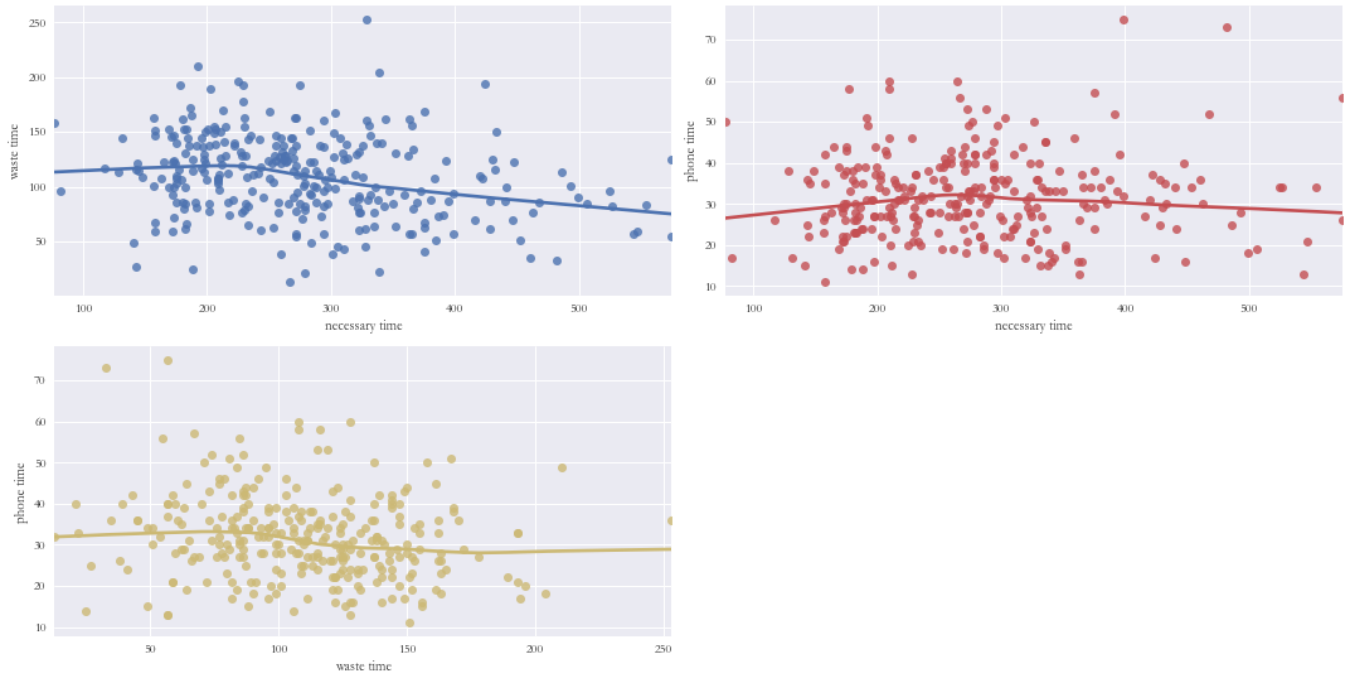
```
    Number of String-Boolean Columns =  0
    Number of Numeric-Boolean Columns =  0
    Number of Discrete String Columns =  1
    Number of NLP String Columns =  0
    Number of Date Time Columns =  0
    Number of ID Columns =  1
    Number of Columns to Delete =  0
No of columns classified 7 does not match 8 total cols. Continuing...
 Missing columns = ['week_order']
       2 variables removed since they were ID or low-information variables

################ Regression VISUALIZATION Started ####################
Using Lowess Smoothing. This might take a few minutes for large data sets...
```

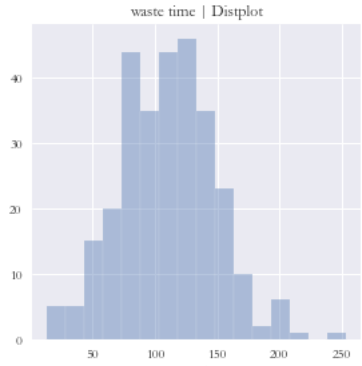
Scatter Plot of each Continuous Variable vs Target

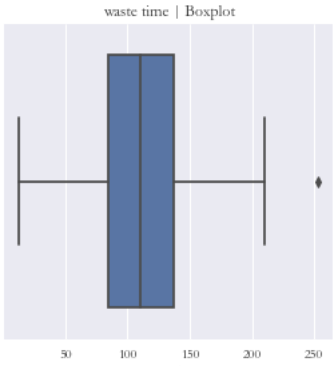


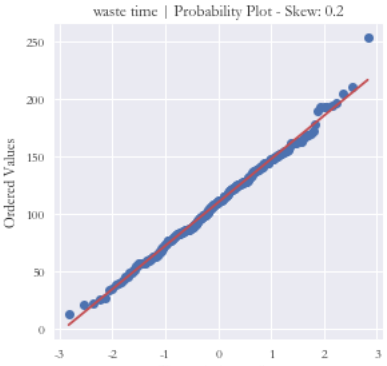

```
Using Lowess Smoothing. This might take a few minutes for large data sets...
Number of All Scatter Plots = 6
```

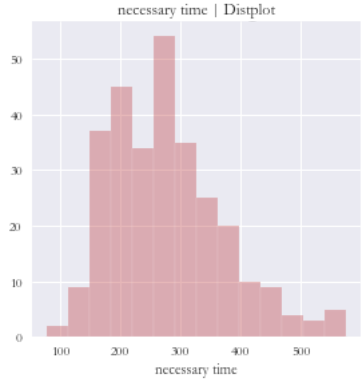
Pair-wise Scatter Plot of all Continuous Variables

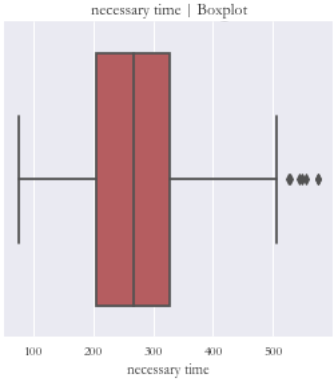


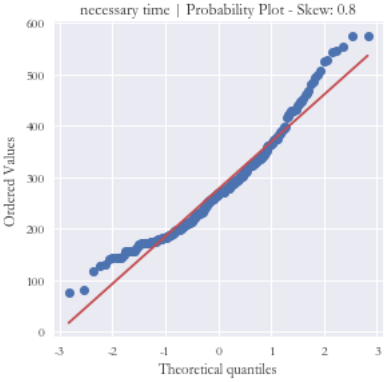

waste time | Distplot
waste time

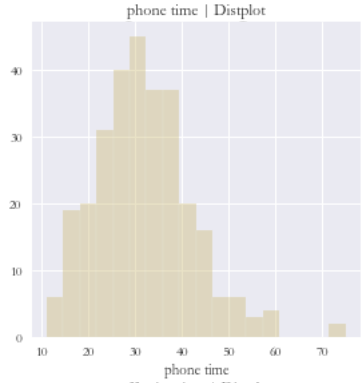

waste time | Boxplot
waste time

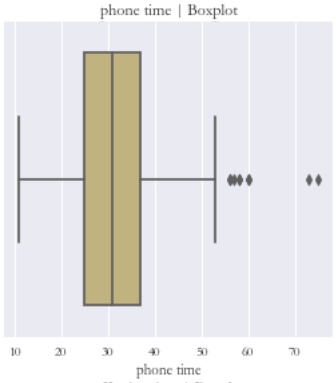

waste time | Probability Plot - Skew: 0.2
Ordered Values
Theoretical quantiles

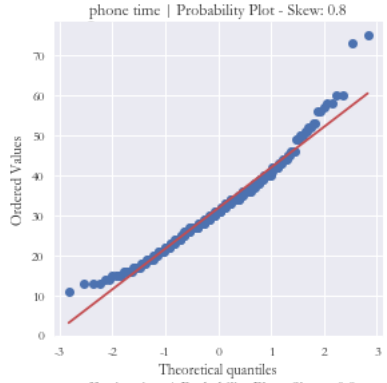

necessary time | Distplot
necessary time

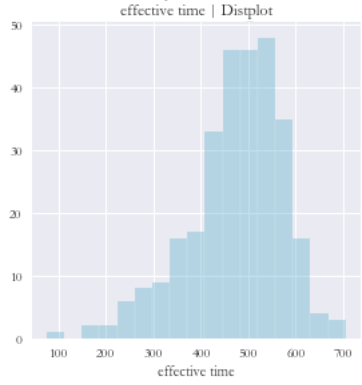

necessary time | Boxplot
necessary time

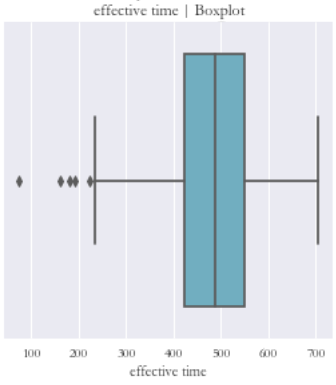

necessary time | Probability Plot - Skew: 0.8
Ordered Values
Theoretical quantiles

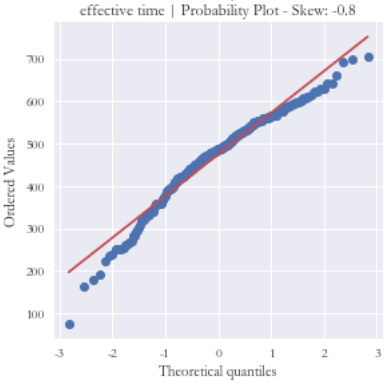

phone time | Distplot
phone time


phone time | Boxplot
phone time

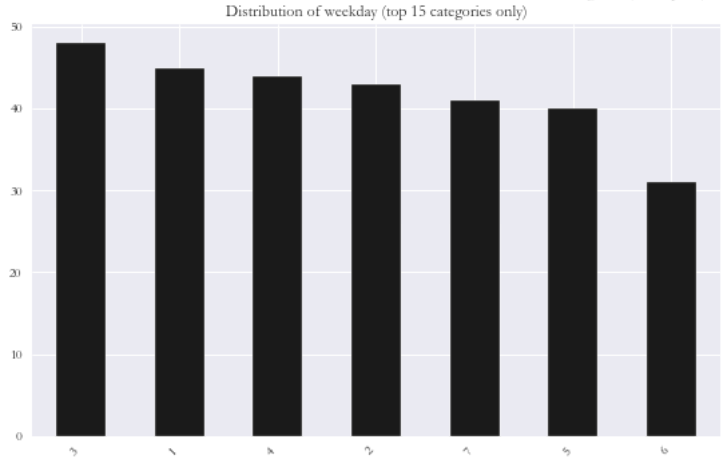

phone time | Probability Plot - Skew: 0.8
Ordered Values
Theoretical quantiles


effective time | Distplot
effective time

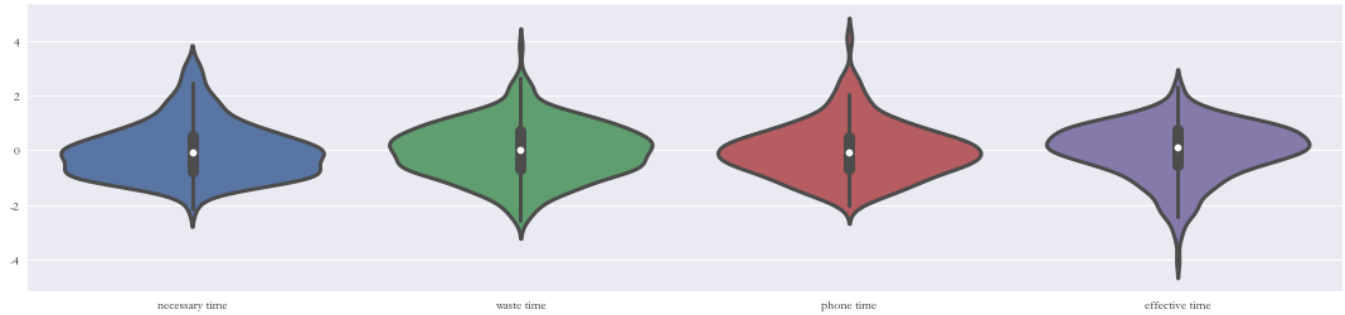

effective time | Boxplot
effective time


effective time | Probability Plot - Skew: -0.8
Ordered Values
Theoretical quantiles

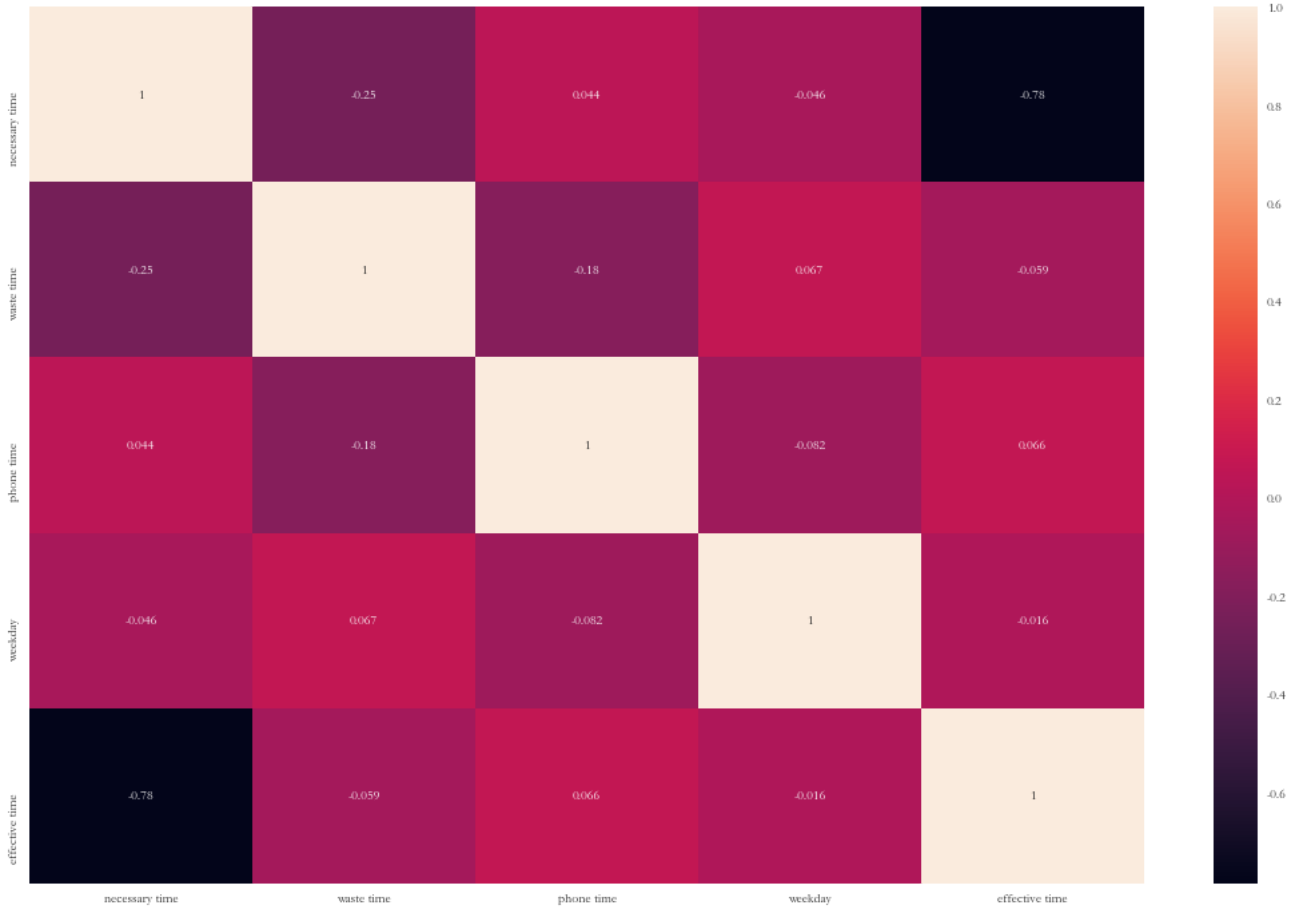
Histograms (KDE plots) of all Continuous Variables

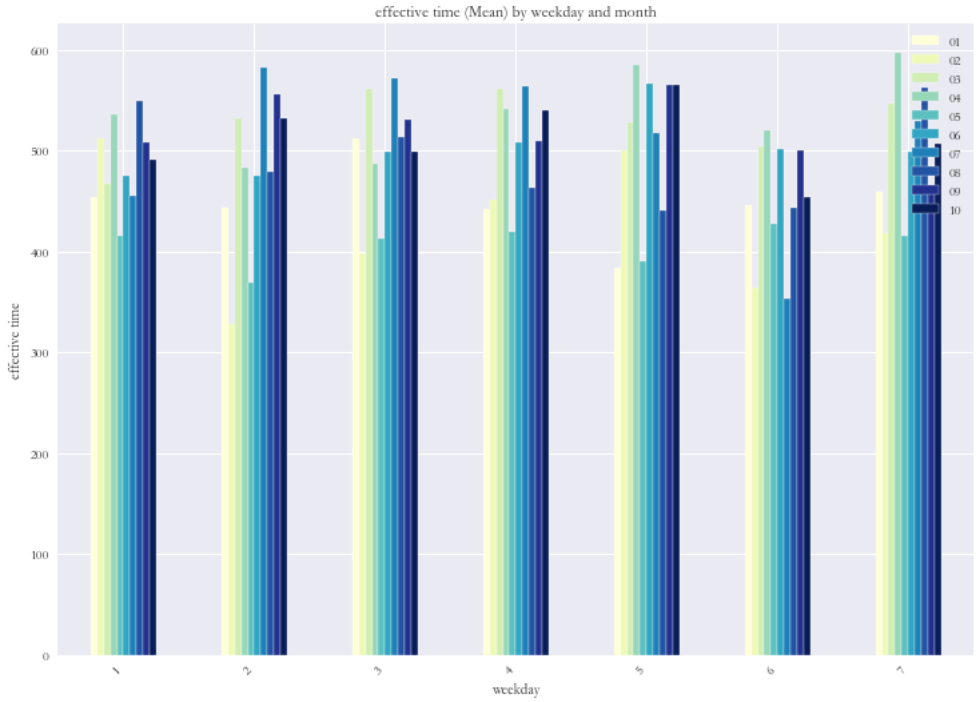


Violin Plot of all Continuous Variables

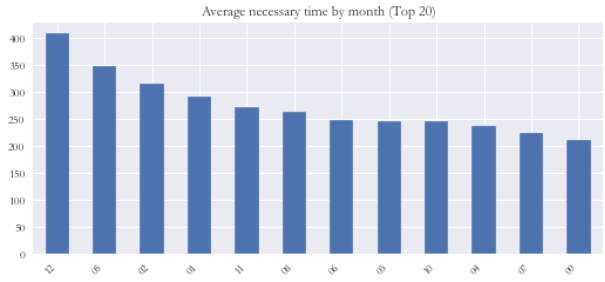


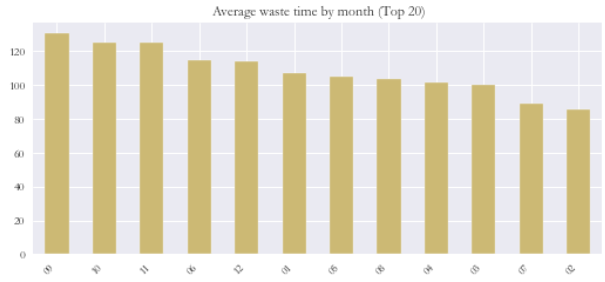
Heatmap of all Continuous Variables including target = effective time

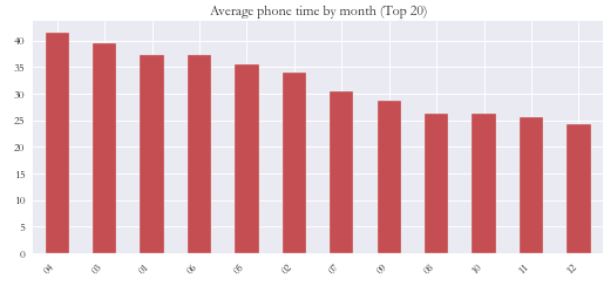


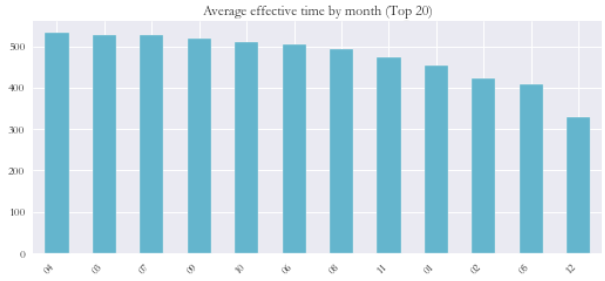

effective time (Mean) by weekday and month
effective time
weekday
0
100
200
300
400
500
600
1
2
3
4
5
6
7
01
02
03
04
05
06
07
08
09
10

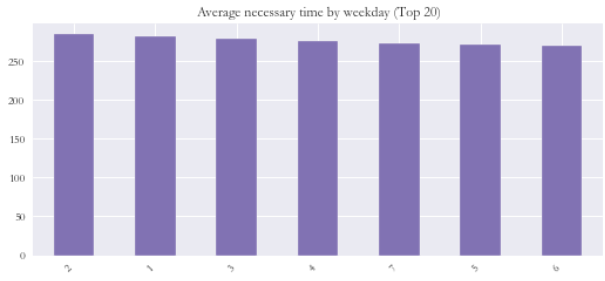
Bar plots for each Continuous by each Categorical variable

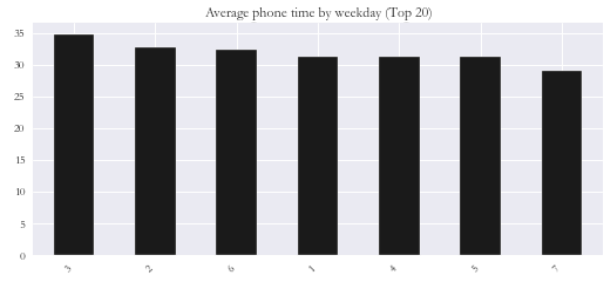


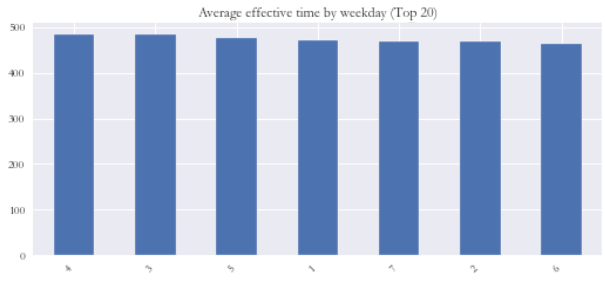













```
All Plots done
Time to run AutoViz (in seconds) = 3.453

 ###################### VISUALIZATION Completed ########################
```

3.数据分析

3.1 方差分析每月或是每年的事项是否有差异

3.1.1 原始事项时长数据的方差分析

```
# 原始事项数据的方差分析
part = '3.1.1'

def vr_event(data, num, cate):
    formula = f'{num}~C({cate})'
    anova_re_2019 = anova_lm(ols(formula, data=data[[num, cate]]).fit())
    res = anova_re_2019.loc[f'C({cate})', 'PR(>F)']
    return res
```

```
# 事项时长数据
vr_duration_table = pd.DataFrame()
for event in tqdm(df['event'].unique()):
    for cate in cate_vars:
        if cate not in ['event', 'attr']:
            temp = df[df['event'] == event].copy()
            temp[cate] = temp[cate].astype(str)
            res = vr_event(temp, 'duration', cate)
            vr_duration_table.loc[event, cate] = res

# 该表格是各个事项的时长数据关于各分类变量的方差分析结果
vr_duration_table
```

```
100%|██████████| 243/243 [00:14<00:00, 16.66it/s]
```

```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

	year	month	day	weekday	mid_hour	day_period	week_order	duration_attr
托福写作与听力	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
python	0.008460	2.287125e-02	0.739882	8.399767e-01	5.757493e-11	1.195543e-08	1.238277e-01	9.591669e-44
快递	0.010206	3.445722e-01	0.730032	1.954784e-01	3.759866e-01	1.737118e-01	4.138766e-01	2.937760e-09
搜索术	NaN	1.000000e+00	NaN	5.792850e-15	NaN	NaN	1.000000e+00	1.113672e-01
午休	0.000013	4.856760e-08	0.864952	2.972783e-01	2.546873e-05	1.956907e-01	6.523863e-15	9.438105e-45
...	...	...	...	...	...	...	...	...
大数据	NaN	NaN	0.038696	6.211273e-01	5.265412e-01	6.896051e-01	5.445203e-01	4.238756e-04
线上研讨	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
做视频	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
报名考试	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN
外出游玩	NaN	NaN	NaN	NaN	NaN	NaN	NaN	NaN

243 rows × 8 columns

```
# 选取那些在不同分类情况下时长有显著差异的数据
vr_duration_table.to_excel('事项时长数据的方差分析结果.xlsx', index=True)
significant_event = vr_duration_table[vr_duration_table > .05].dropna(axis=0, how='all')
significant_event
```

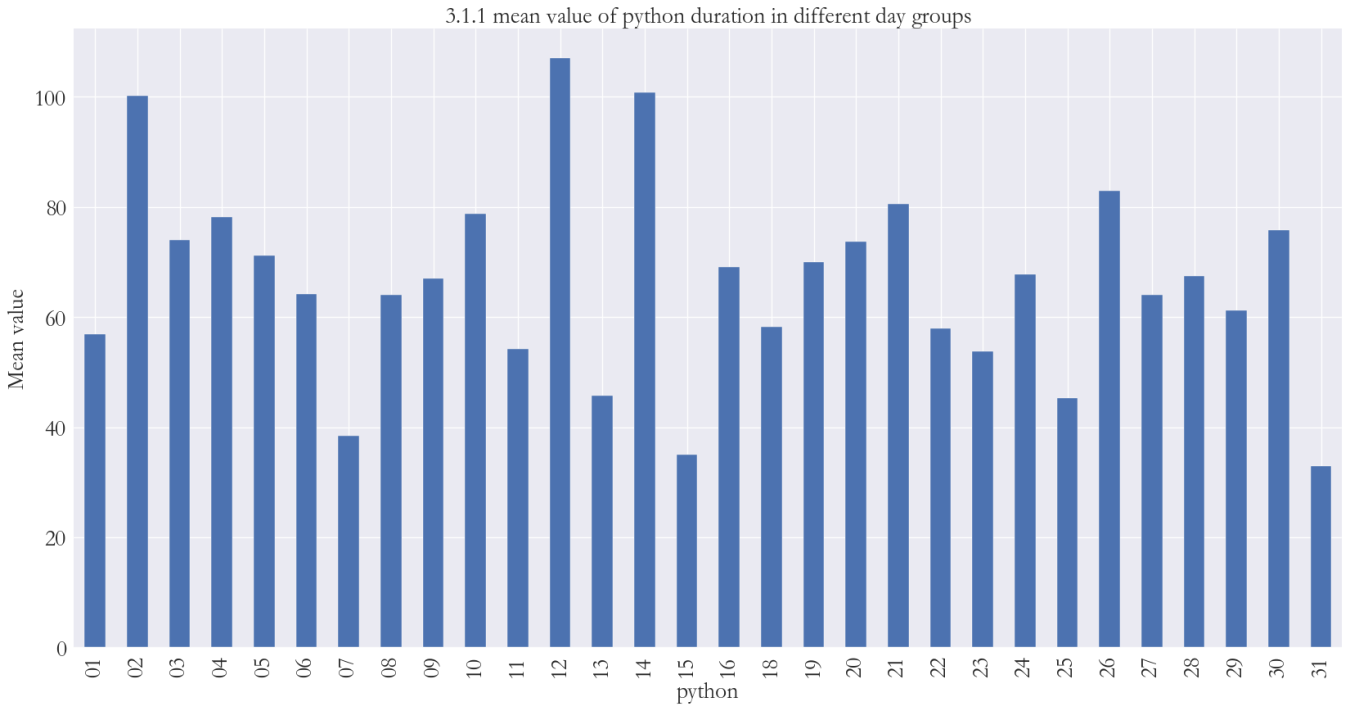
```
.dataframe tbody tr th {
    vertical-align: top;
}

.dataframe thead th {
    text-align: right;
}
```

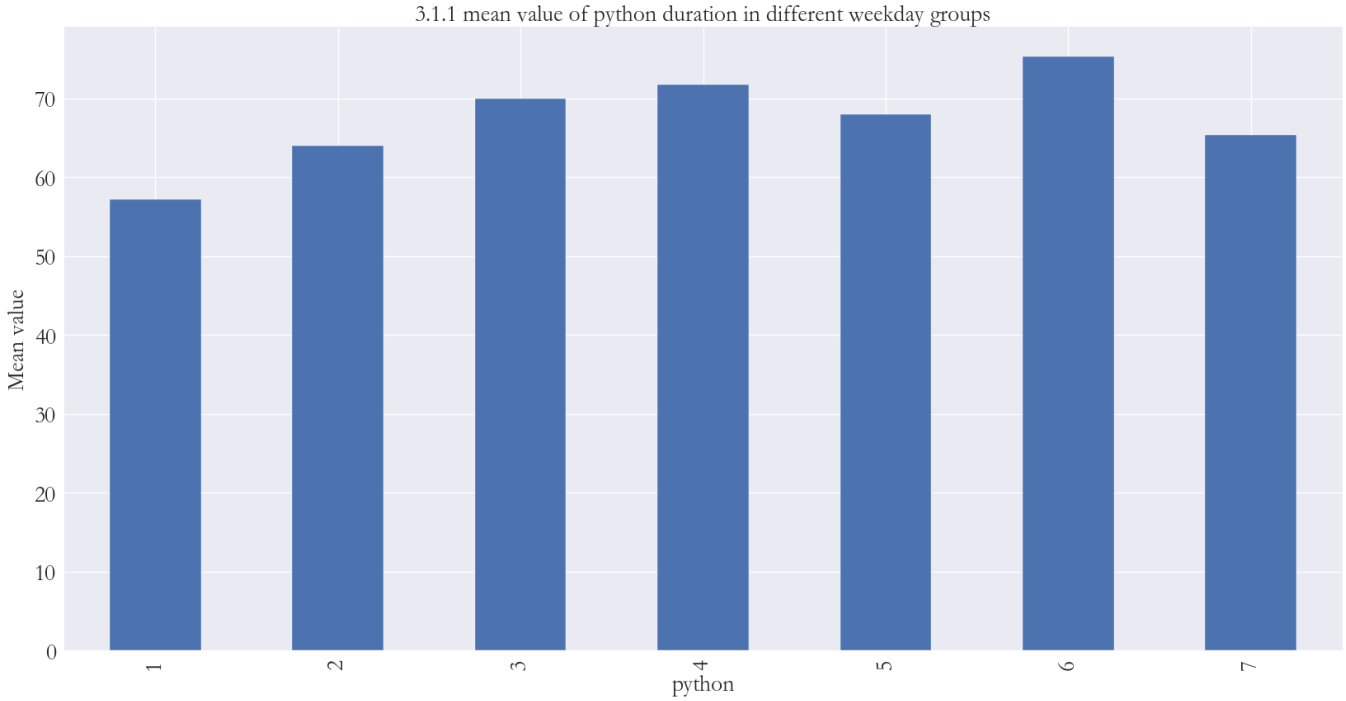
	year	month	day	weekday	mid_hour	day_period	week_order	duration_attr
python	NaN	NaN	0.739882	0.839977	NaN	NaN	0.123828	NaN
快递	NaN	0.344572	0.730032	0.195478	0.375987	0.173712	0.413877	NaN
搜索术	NaN	1.000000	NaN	NaN	NaN	NaN	1.000000	0.111367
午休	NaN	NaN	0.864952	0.297278	NaN	0.195691	NaN	NaN
托福听力	NaN	NaN	0.786872	0.706346	NaN	NaN	NaN	NaN
...	...	...	...	...	...	...	...	...
机器学习	NaN	0.434528	0.665255	0.402006	NaN	NaN	0.141308	NaN
制定计划	NaN	0.544331	NaN	NaN	0.592593	0.592593	NaN	0.074074
配置环境	NaN	NaN	1.000000	1.000000	0.333333	0.333333	NaN	NaN
深度学习	NaN	0.378635	0.731785	0.193446	NaN	0.524222	0.113776	NaN
大数据	NaN	NaN	NaN	0.621127	0.526541	0.689605	0.544520	NaN

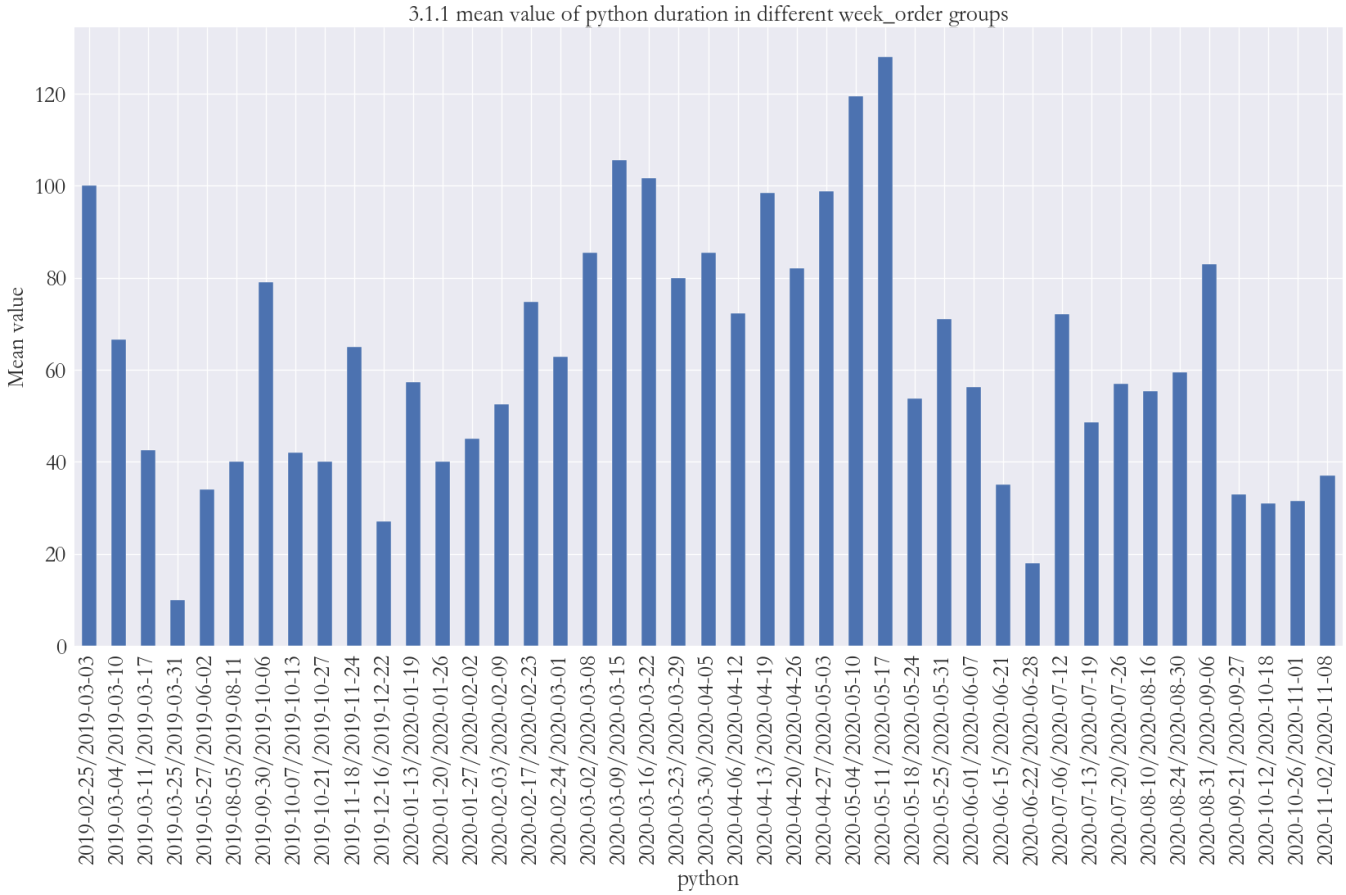
114 rows × 8 columns

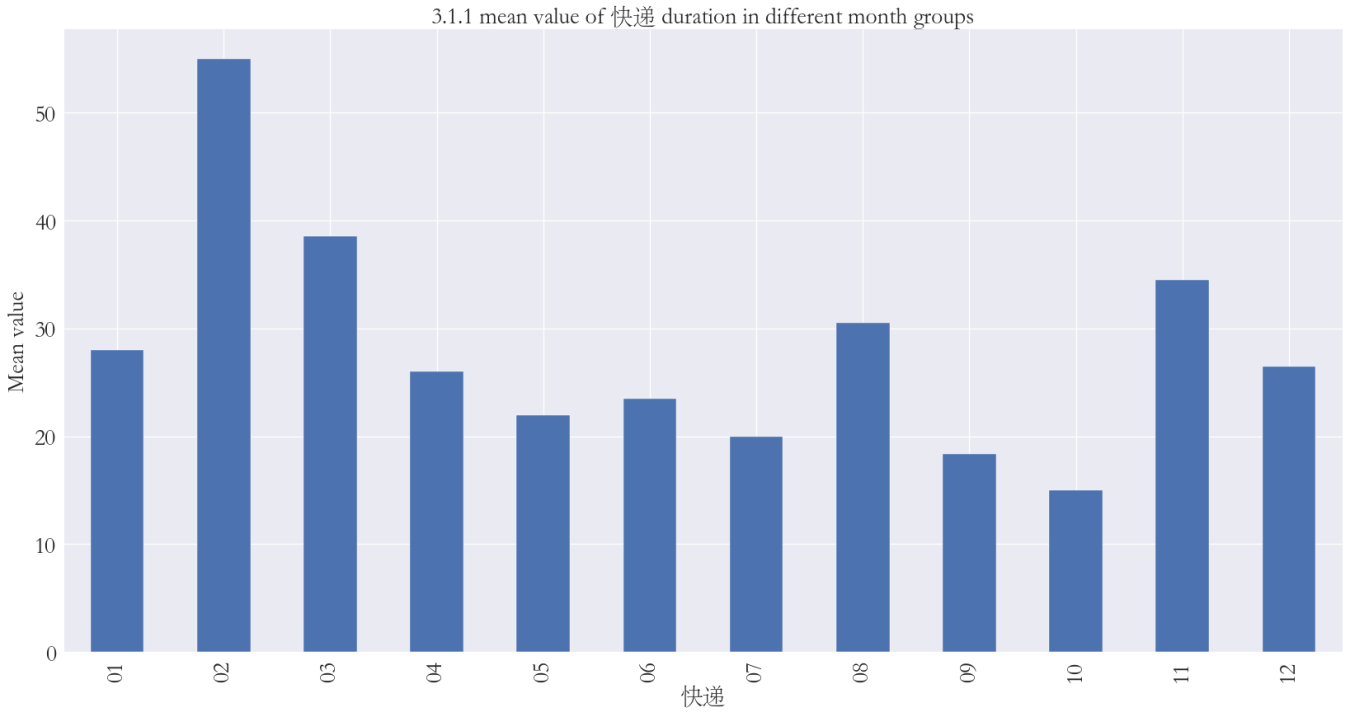
```
# 对于有显著差异的事项 做出在不同分类中均值的bar图
mins = 30  # 剔除掉出现次数少的事项
for e in tqdm(significant_event.index):
    if len(df[df['event'] == e]) >= mins:
        for c in significant_event.columns:
            if significant_event.loc[e, c] >= .05:
                plt.figure(dpi=100)
                df[df['event'] == e].groupby(c)['duration'].agg('mean').plot(
                    kind='bar', figsize=(20, 10)
                )
                title = f'{part} mean value of {e} duration in different {c} groups'
                plt.title(title, fontsize=20)
                plt.xlabel(e, fontsize=20)
                plt.ylabel('Mean value', fontsize=20)
                plt.xticks(fontsize=20)
                plt.yticks(fontsize=20)
                file = os.path.join(path, f'{title}.png')
                plt.savefig(file, dpi=600)
```

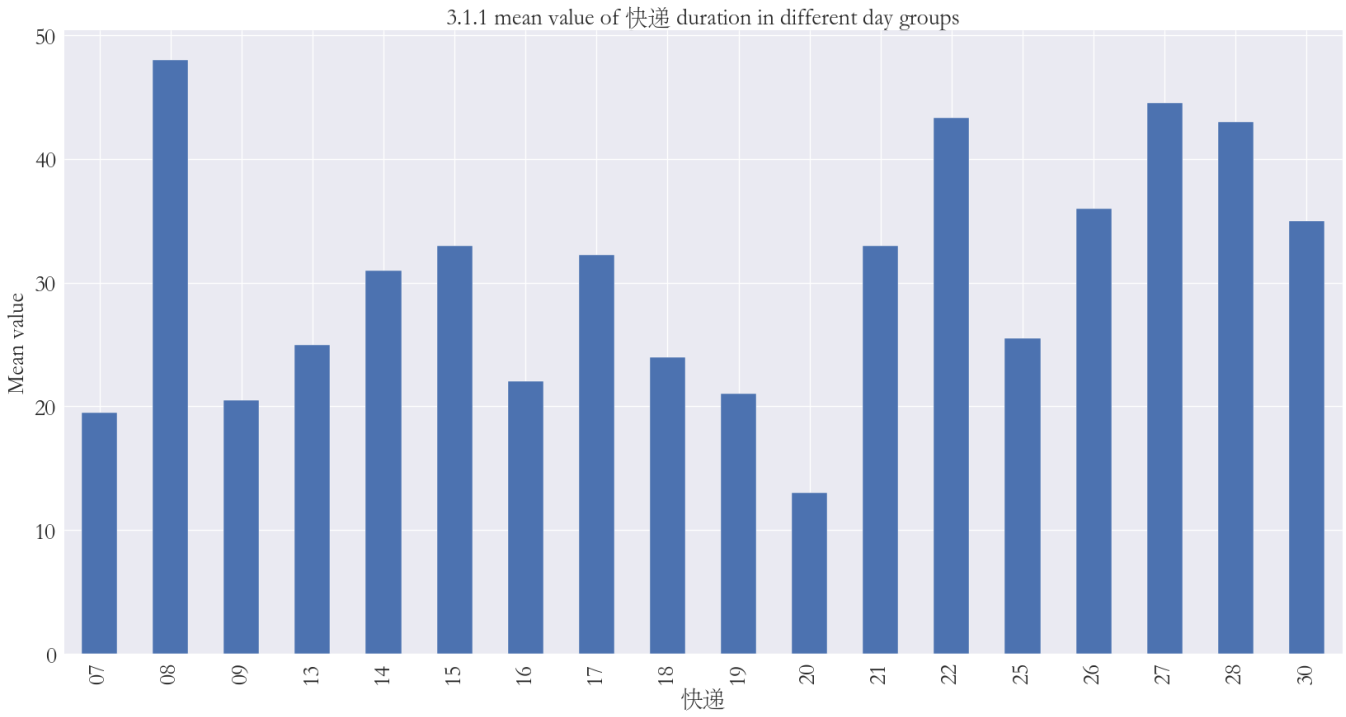
```
100%|██████████| 114/114 [07:23<00:00,  3.89s/it]
```

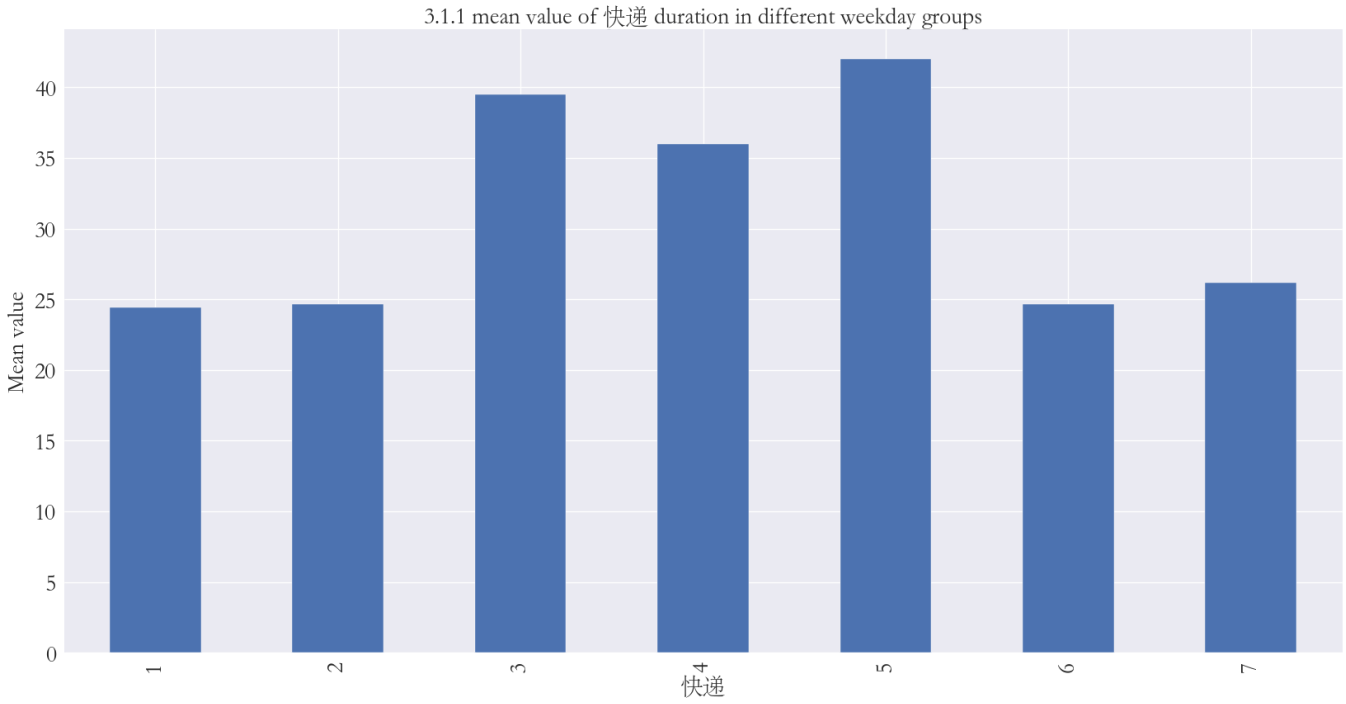


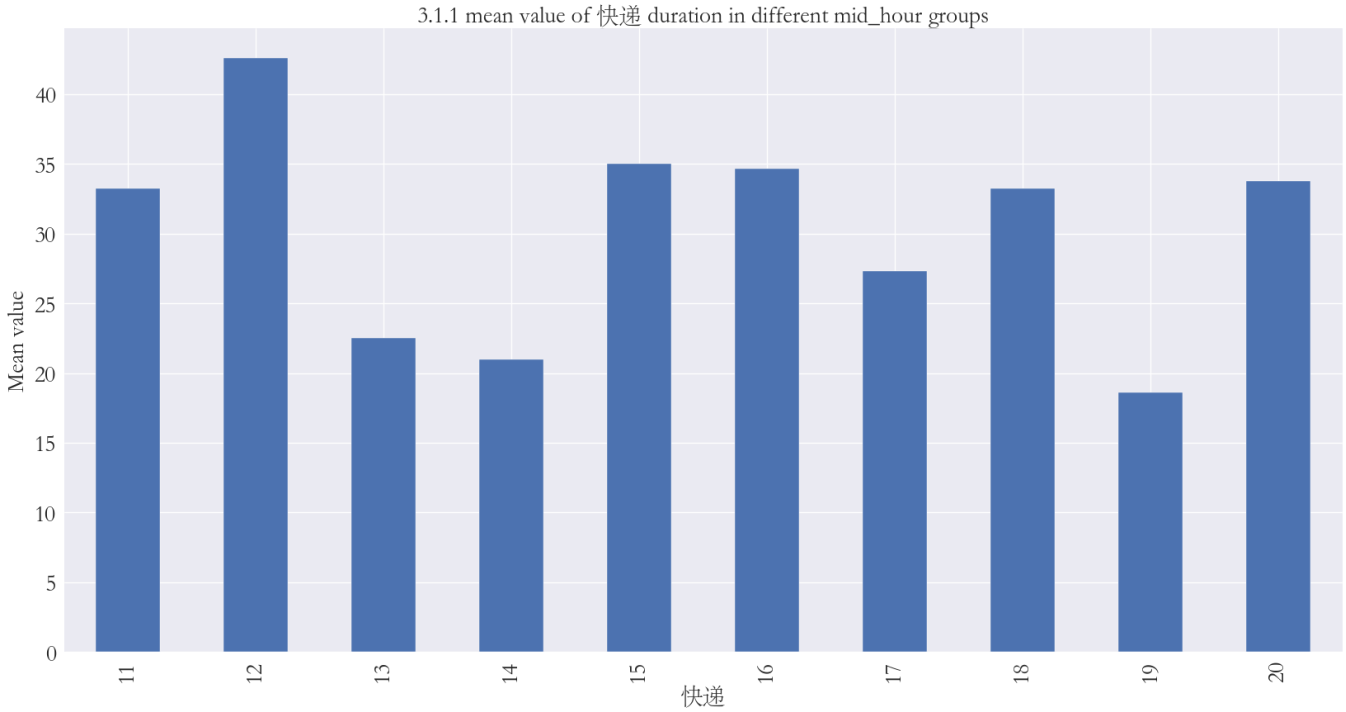

3.1.1 mean value of python duration in different weekday groups
Mean value
0
10
20
30
40
50
60
70
1
2
3
4
5
6
7
python

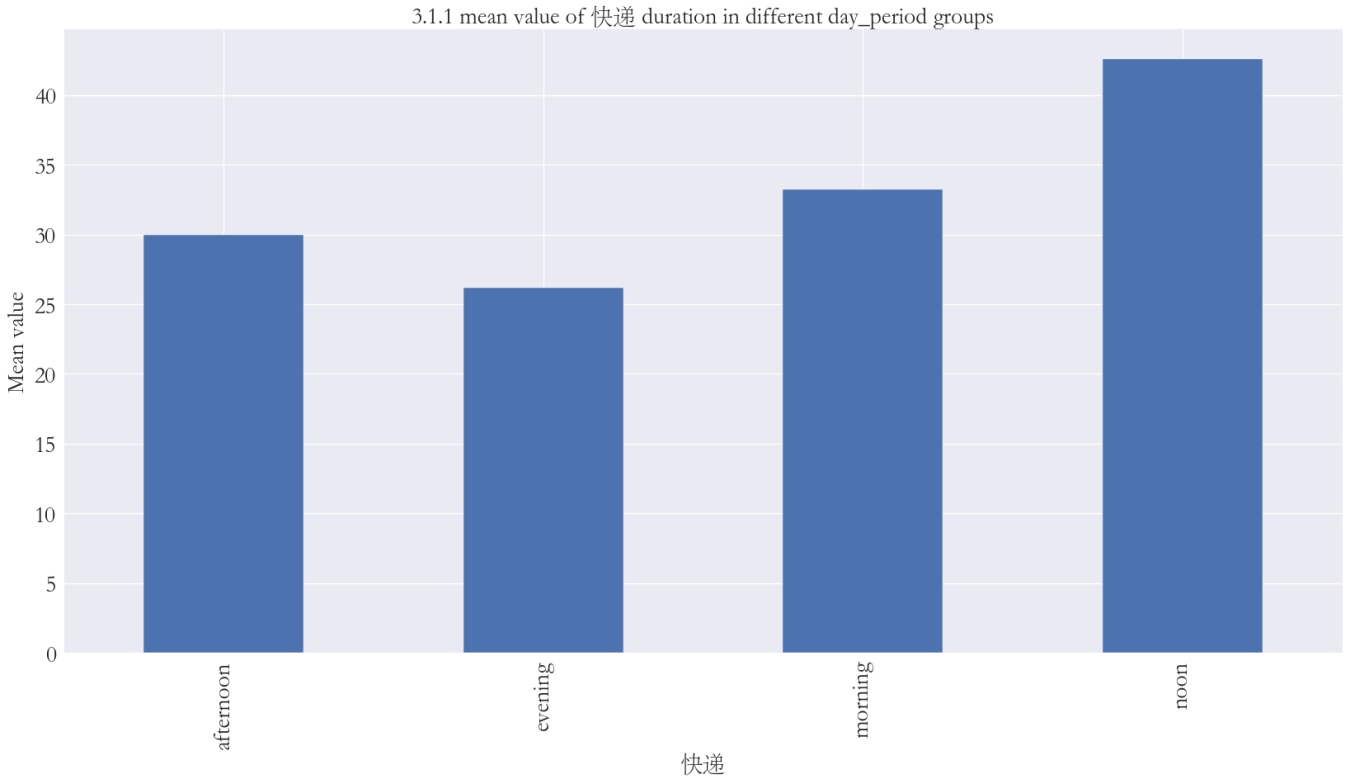

3.1.1 mean value of python duration in different week_order groups
Mean value
0
20
40
60
80
100
120
2019-02-25/2019-03-03
2019-03-04/2019-03-10
2019-03-11/2019-03-17
2019-03-25/2019-03-31
2019-05-27/2019-06-02
2019-08-05/2019-08-11
2019-09-30/2019-10-06
2019-10-07/2019-10-13
2019-10-21/2019-10-27
2019-11-18/2019-11-24
2019-12-16/2019-12-22
2020-01-13/2020-01-19
2020-01-20/2020-01-26
2020-01-27/2020-02-02
2020-02-03/2020-02-09
2020-02-17/2020-02-23
2020-02-24/2020-03-01
2020-03-02/2020-03-08
2020-03-09/2020-03-15
2020-03-16/2020-03-22
2020-03-23/2020-03-29
2020-03-30/2020-04-05
2020-04-06/2020-04-12
2020-04-13/2020-04-19
2020-04-20/2020-04-26
2020-04-27/2020-05-03
2020-05-04/2020-05-10
2020-05-11/2020-05-17
2020-05-18/2020-05-24
2020-05-25/2020-05-31
2020-06-01/2020-06-07
2020-06-15/2020-06-21
2020-06-22/2020-06-28
2020-07-06/2020-07-12
2020-07-13/2020-07-19
2020-07-20/2020-07-26
2020-08-10/2020-08-16
2020-08-24/2020-08-30
2020-08-31/2020-09-06
2020-09-21/2020-09-27
2020-10-12/2020-10-18
2020-10-26/2020-11-01
2020-11-02/2020-11-08
python

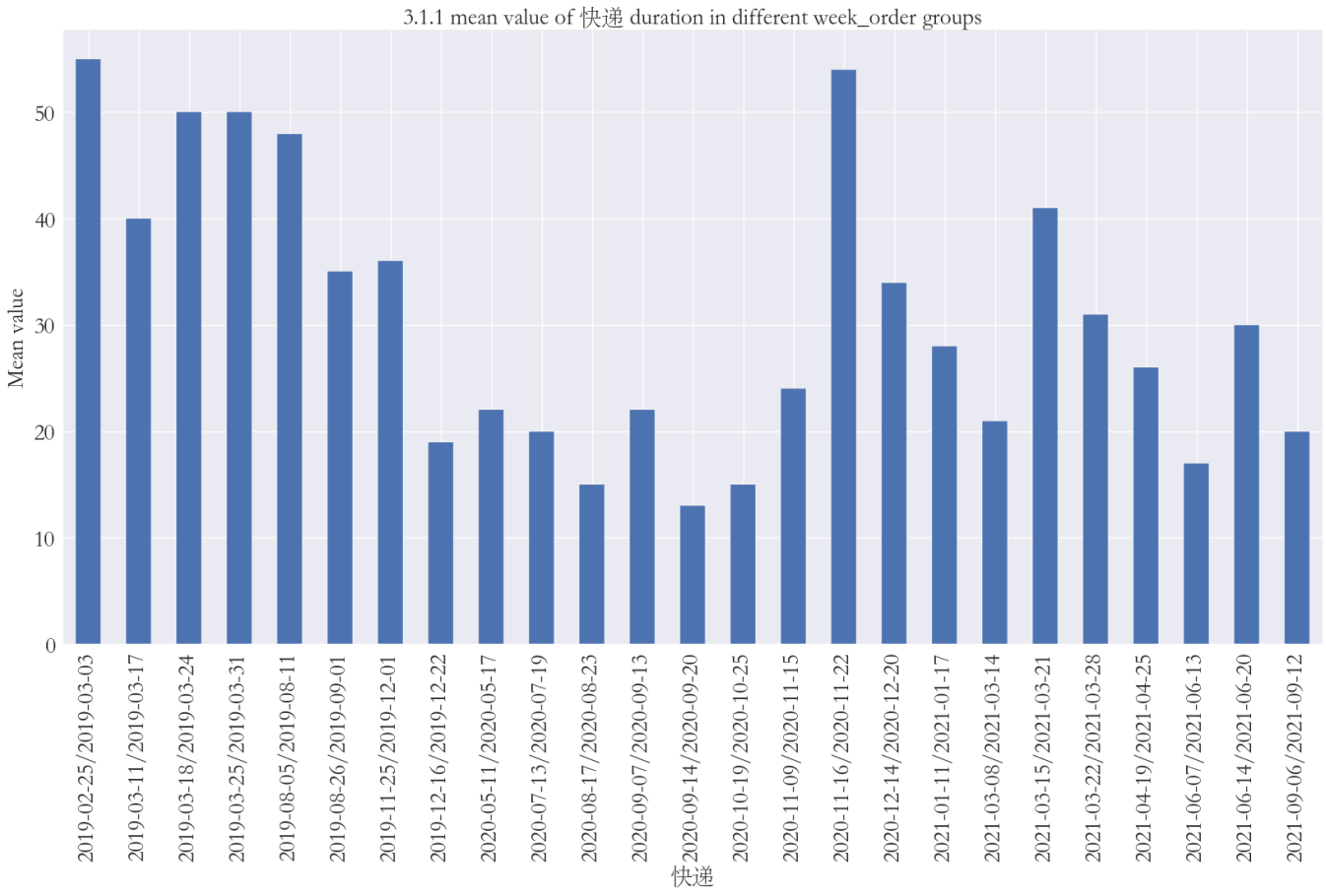

3.1.1 mean value of 快递 duration in different month groups
Mean value
0
10
20
30
40
50
01
02
03
04
05
06
07
08
09
10
11
12
快递

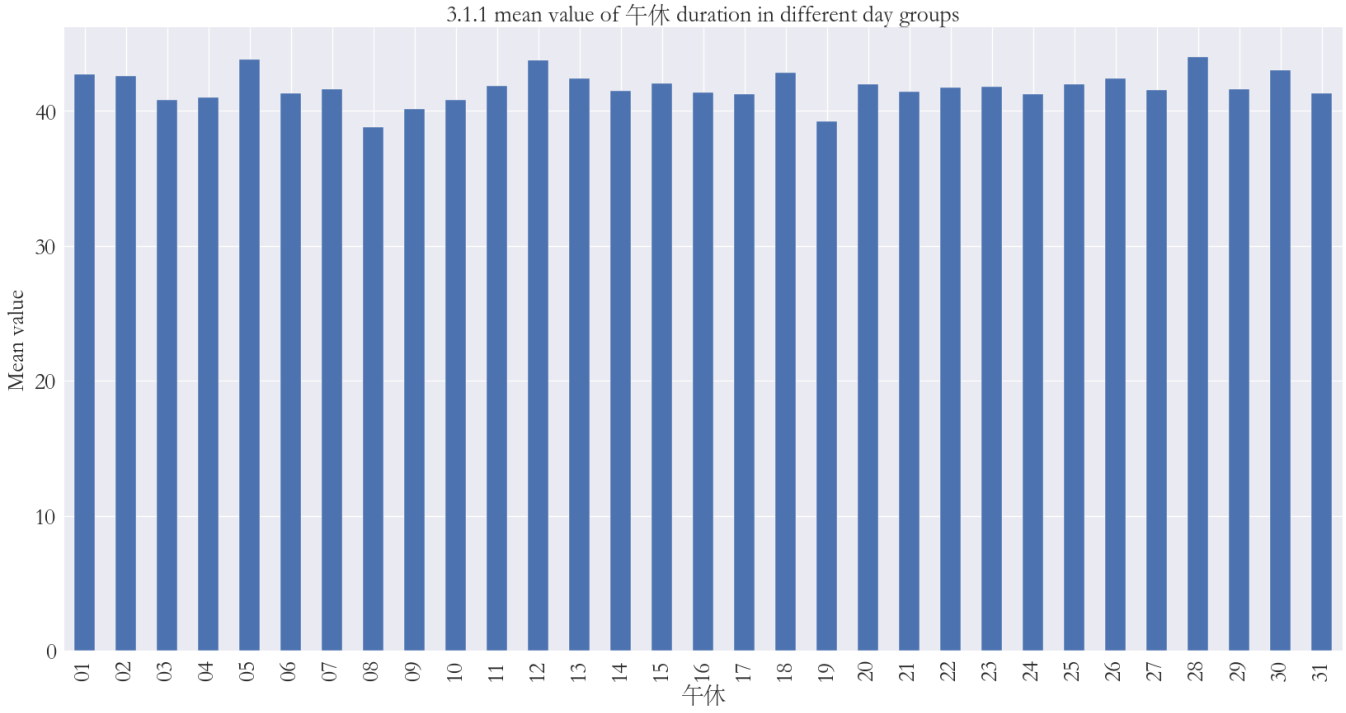

3.1.1 mean value of 快递 duration in different day groups
Mean value
0
10
20
30
40
50
07
08
09
13
14
15
16
17
18
19
20
21
22
25
26
27
28
30
快递

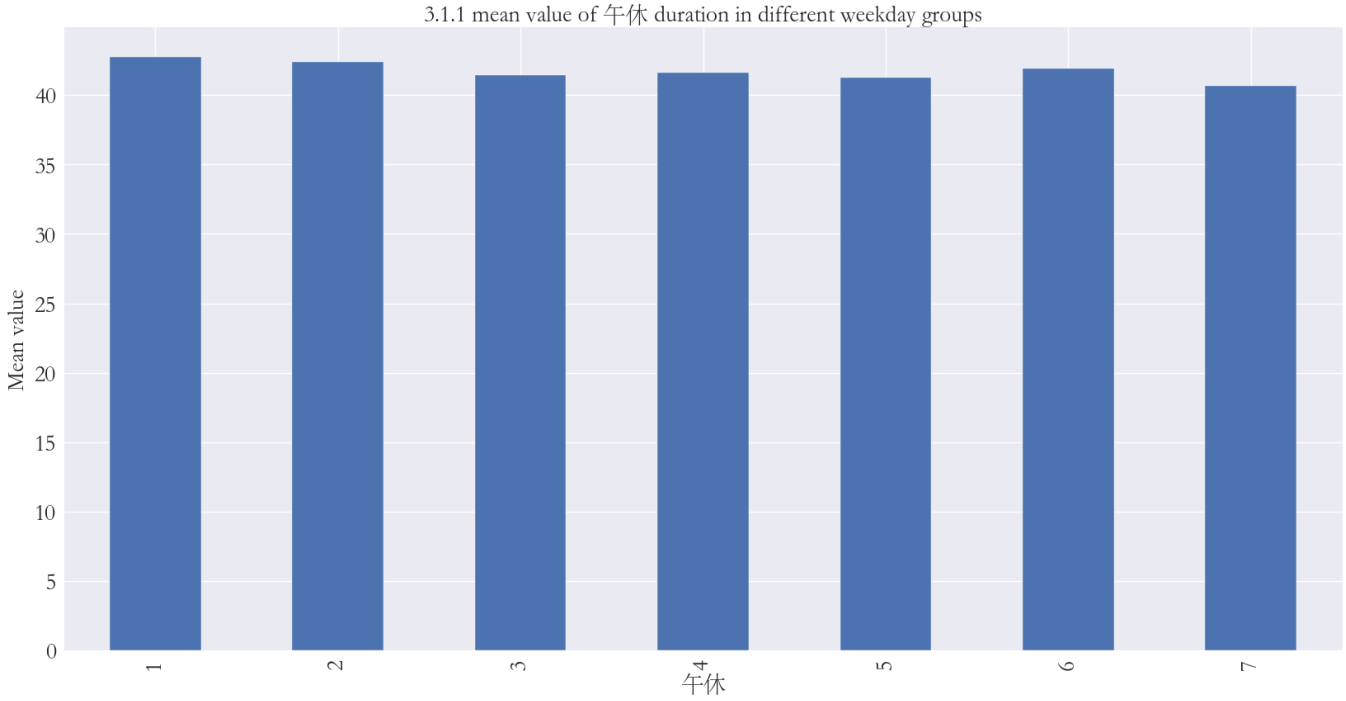

3.1.1 mean value of 快递 duration in different weekday groups
Mean value
0
5
10
15
20
25
30
35
40
1
2
3
4
5
6
7
快递

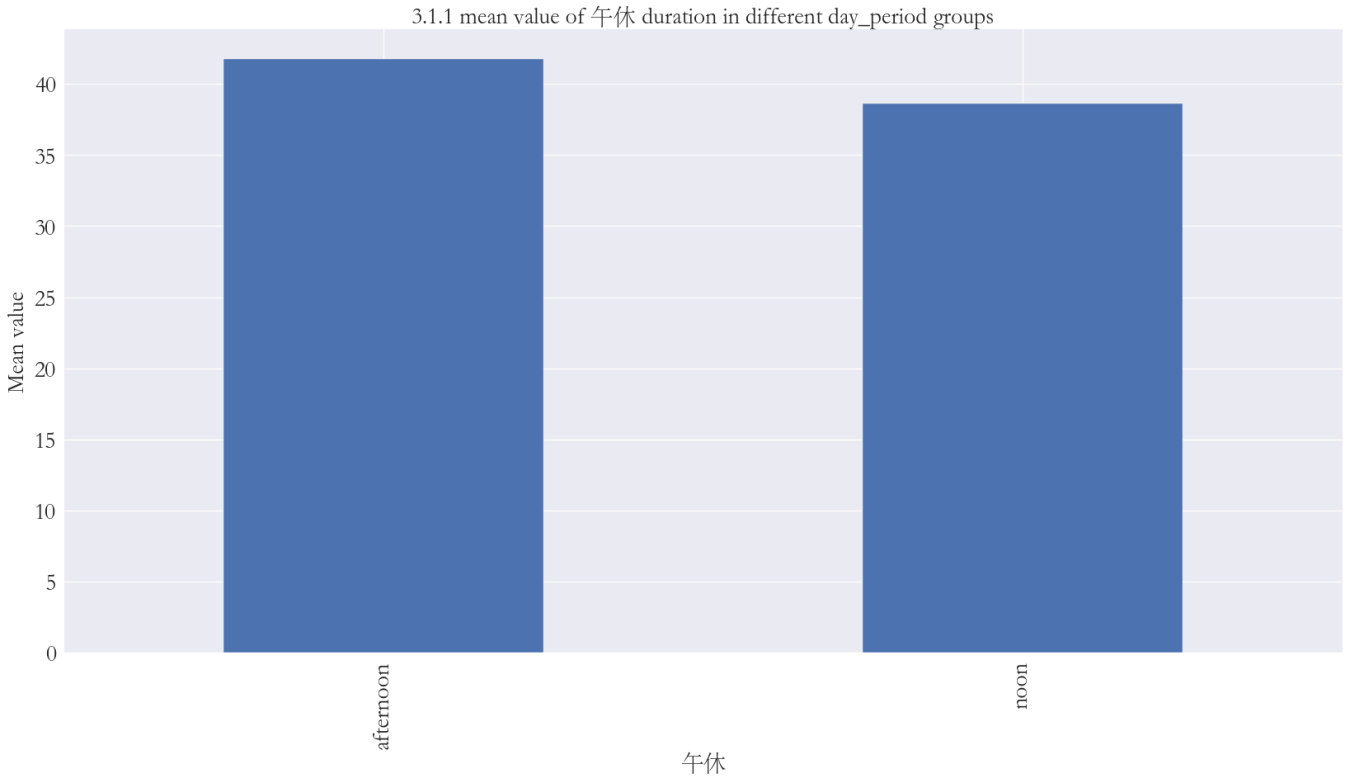

3.1.1 mean value of 快递 duration in different mid_hour groups
Mean value
0
5
10
15
20
25
30
35
40
11
12
13
14
15
16
17
18
19
20
快递

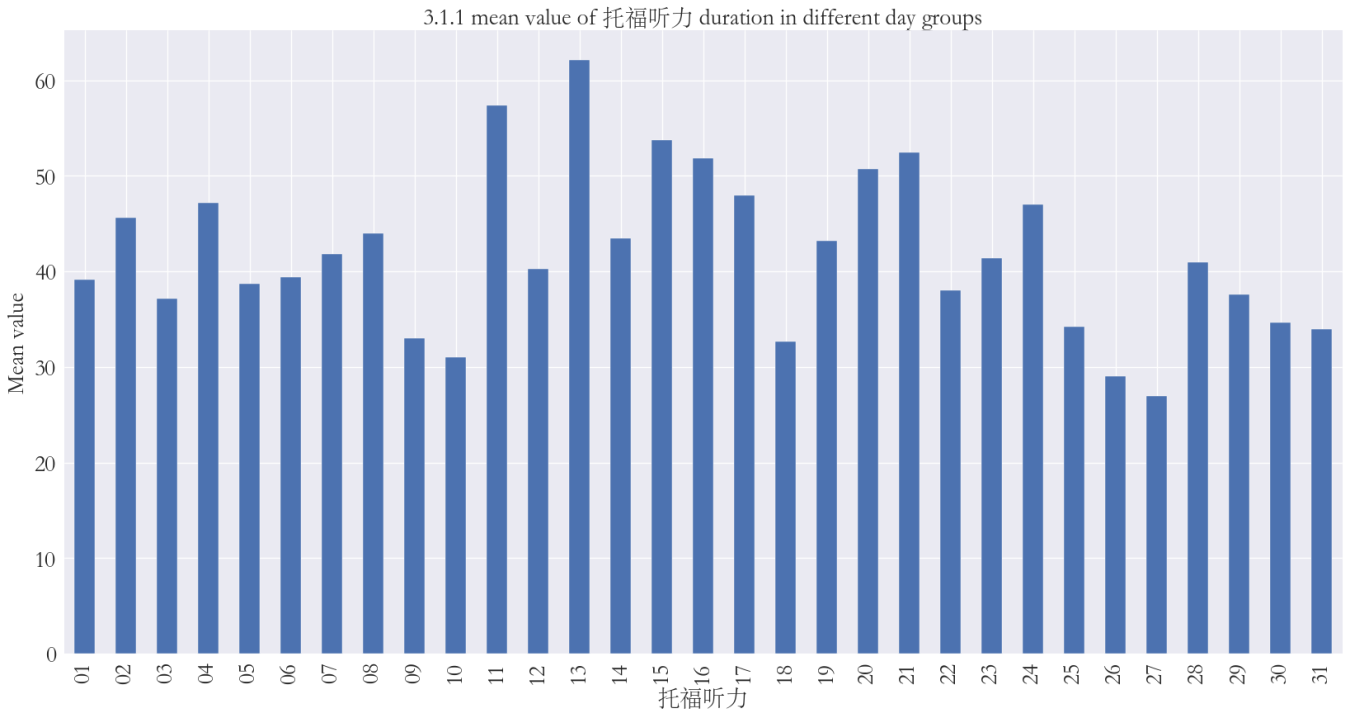

3.1.1 mean value of 快递 duration in different day_period groups
Mean value
0
5
10
15
20
25
30
35
40
afternoon
evening
morning
noon
快递

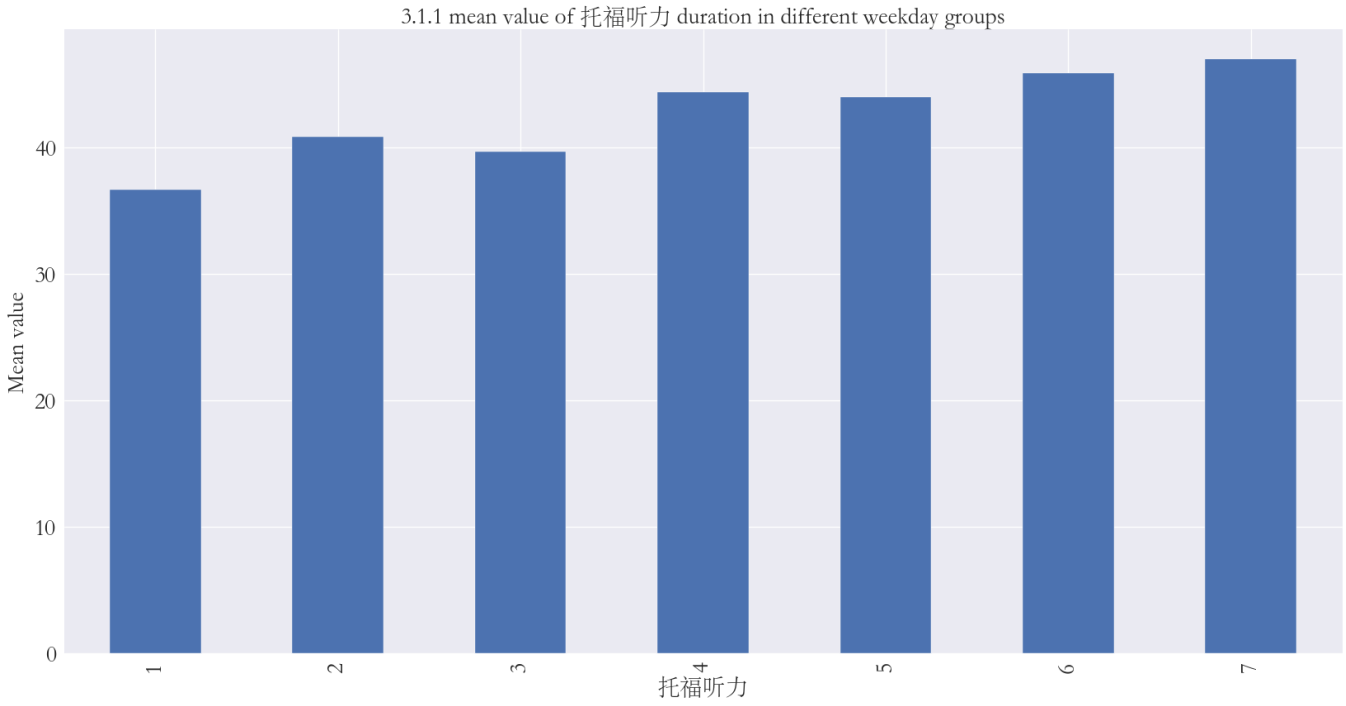

3.1.1 mean value of 快递 duration in different week_order groups
Mean value
0
10
20
30
40
50
2019-02-25/2019-03-03
2019-03-11/2019-03-17
2019-03-18/2019-03-24
2019-03-25/2019-03-31
2019-08-05/2019-08-11
2019-08-26/2019-09-01
2019-11-25/2019-12-01
2019-12-16/2019-12-22
2020-05-11/2020-05-17
2020-07-13/2020-07-19
2020-08-17/2020-08-23
2020-09-07/2020-09-13
2020-09-14/2020-09-20
2020-10-19/2020-10-25
2020-11-09/2020-11-15
2020-11-16/2020-11-22
2020-12-14/2020-12-20
2021-01-11/2021-01-17
2021-03-08/2021-03-14
2021-03-15/2021-03-21
2021-03-22/2021-03-28
2021-04-19/2021-04-25
2021-06-07/2021-06-13
2021-06-14/2021-06-20
2021-09-06/2021-09-12
快递

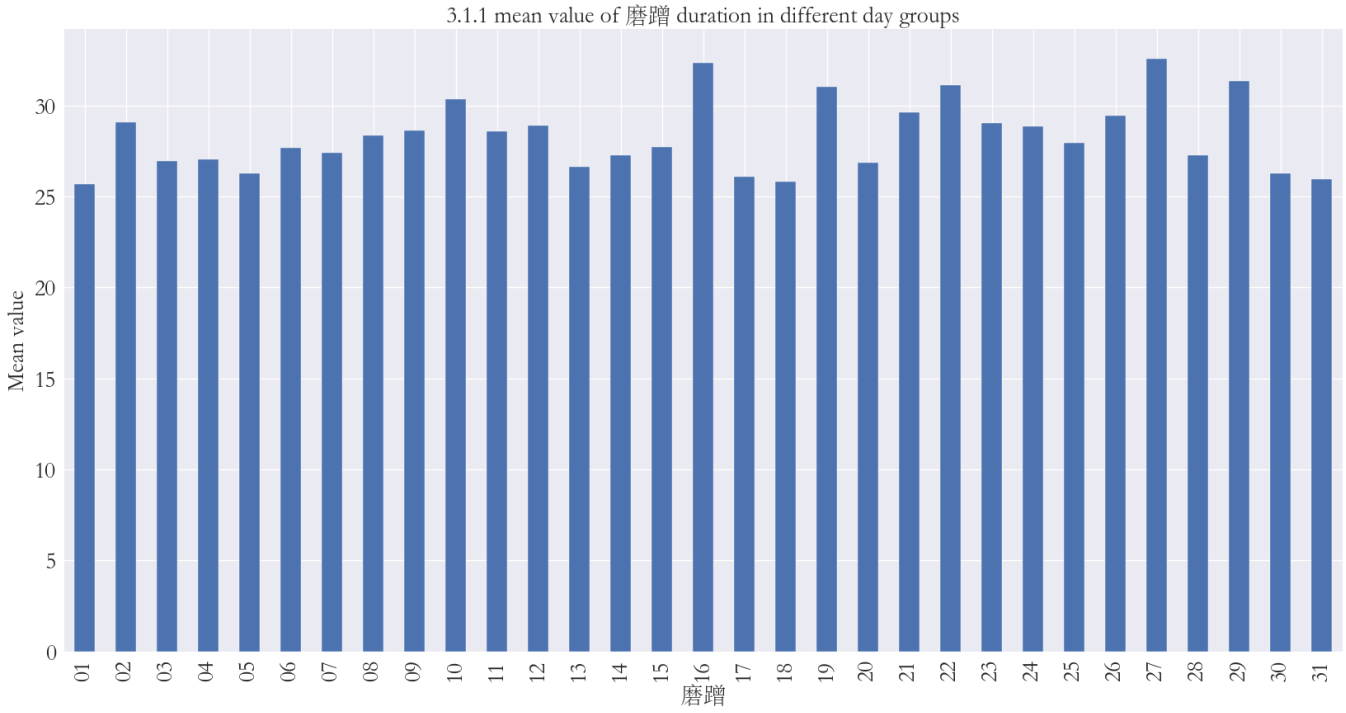

3.1.1 mean value of 午休 duration in different day groups
Mean value
0
10
20
30
40
01
02
03
04
05
06
07
08
09
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
午休

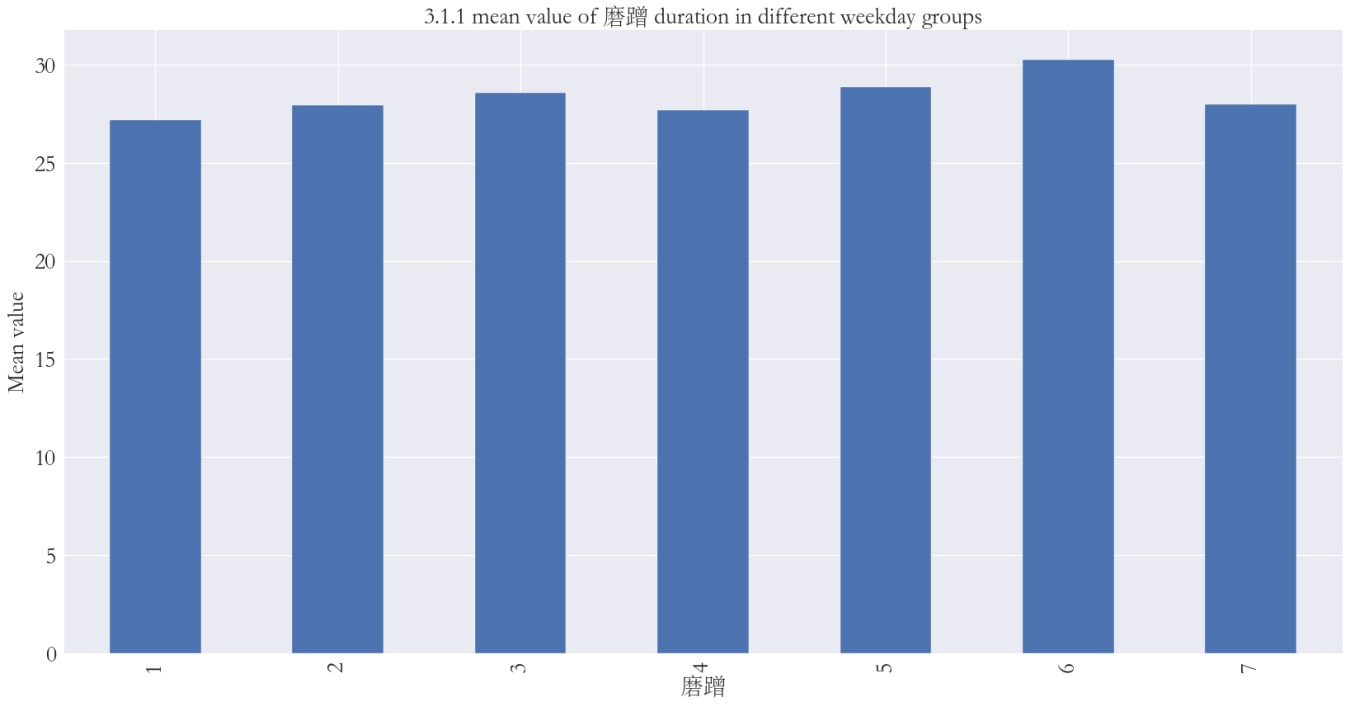

3.1.1 mean value of 午休 duration in different weekday groups
Mean value
0
5
10
15
20
25
30
35
40
1
2
3
4
5
6
7
午休

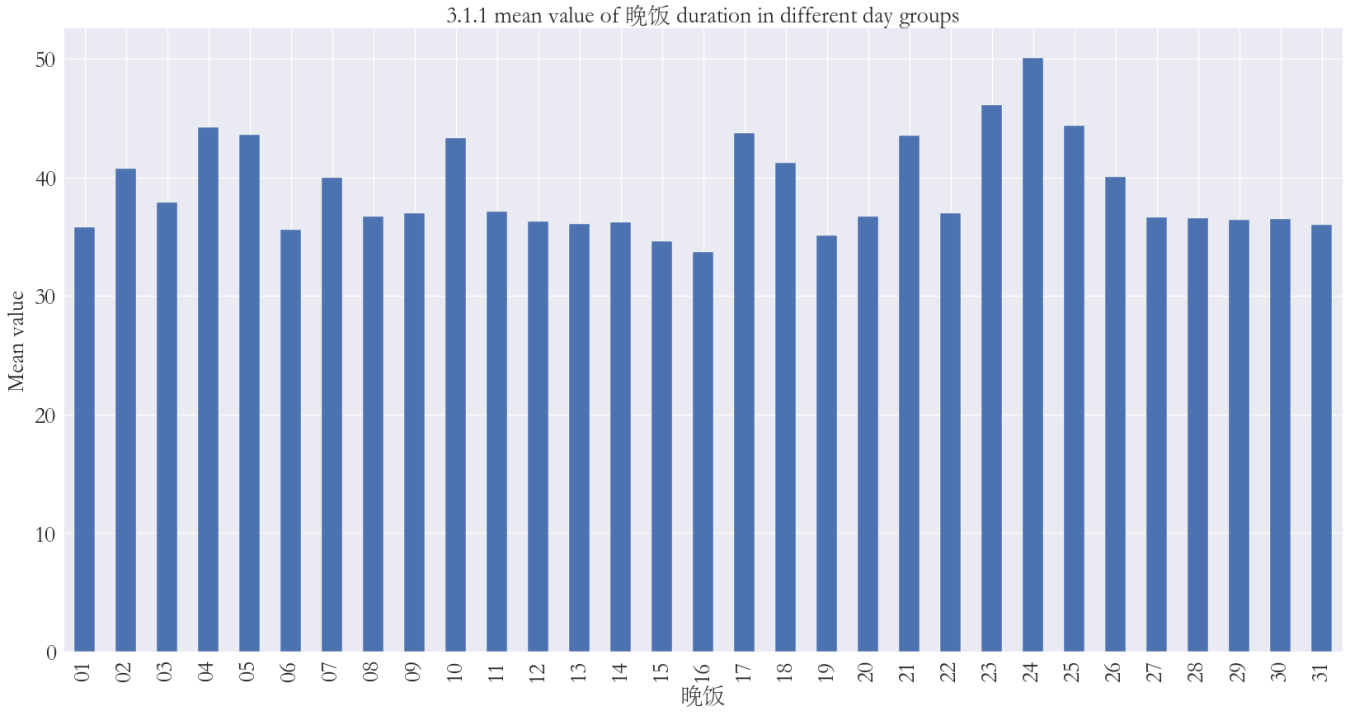

3.1.1 mean value of 午休 duration in different day_period groups
Mean value
0
5
10
15
20
25
30
35
40
afternoon
noon
午休

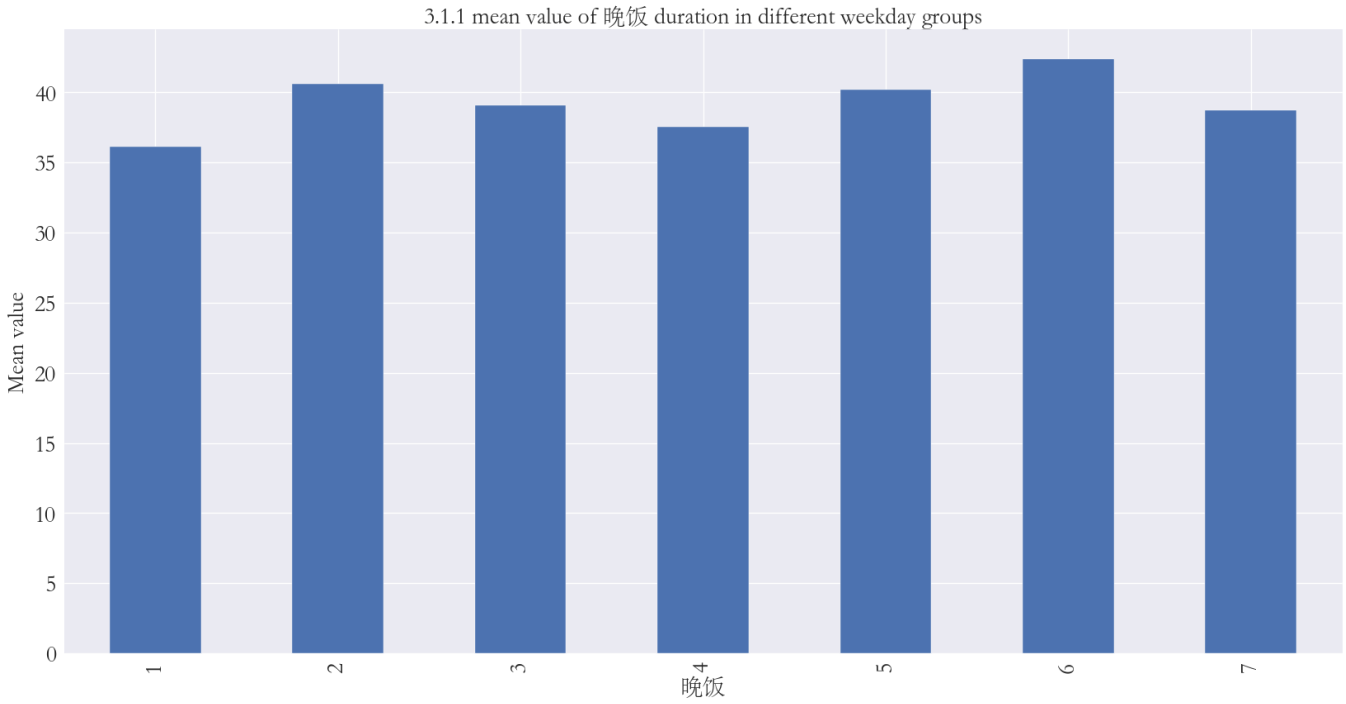

3.1.1 mean value of 托福听力 duration in different day groups
Mean value
0
10
20
30
40
50
60
01
02
03
04
05
06
07
08
09
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
托福听力

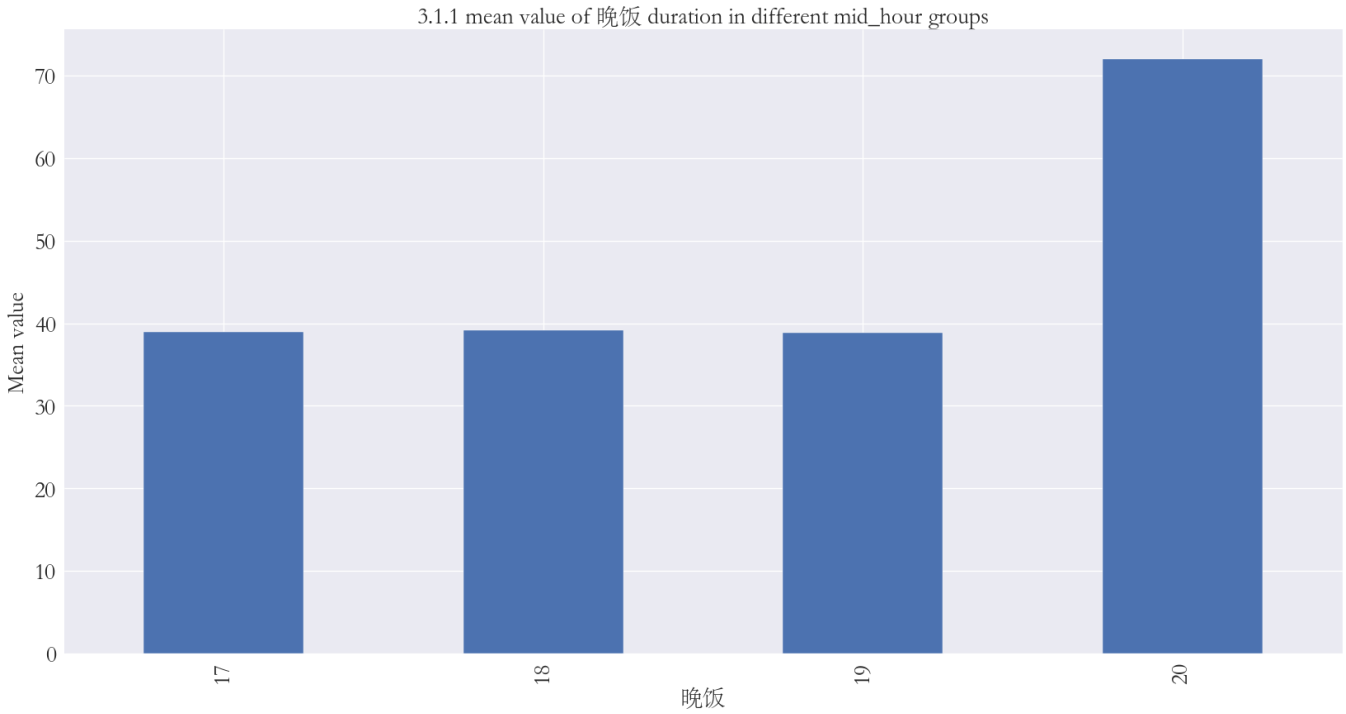

3.1.1 mean value of 托福听力 duration in different weekday groups
Mean value
0
10
20
30
40
1
2
3
4
5
6
7
托福听力

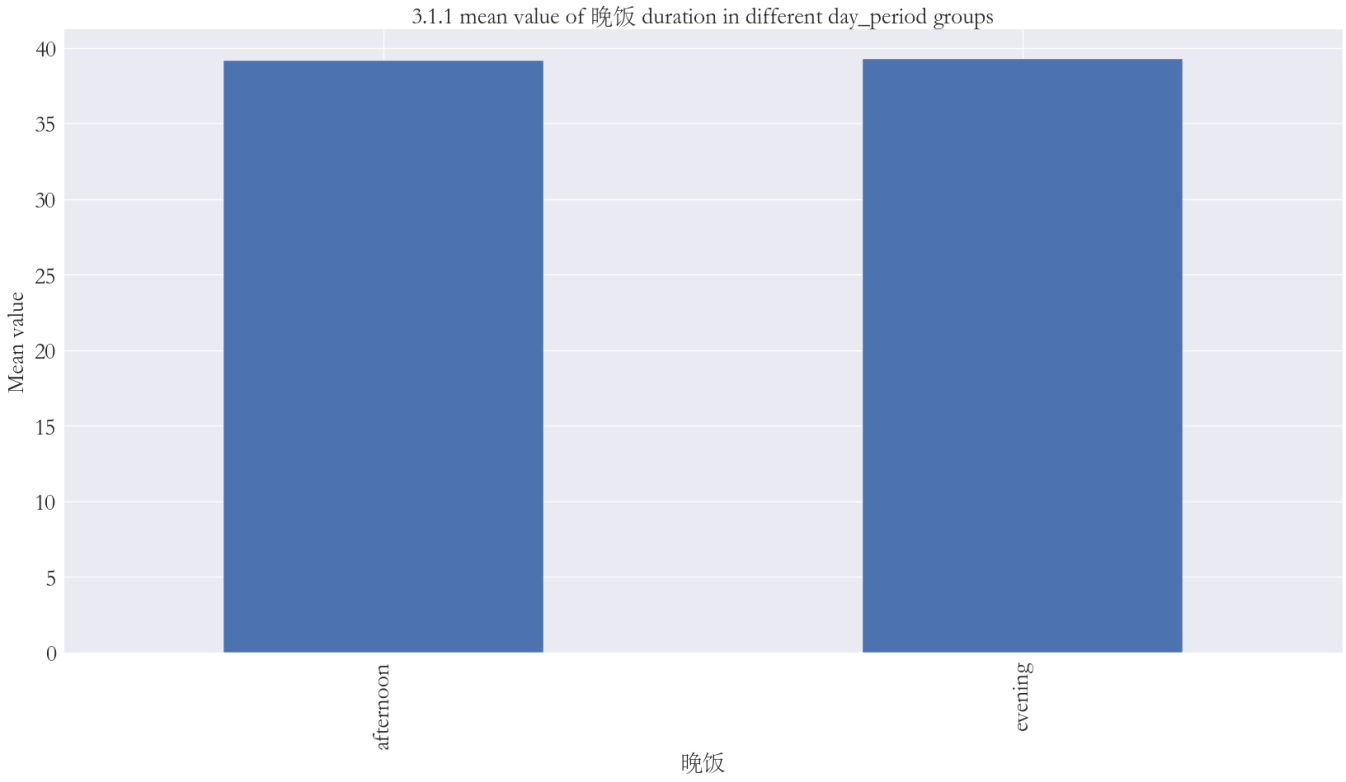

3.1.1 mean value of 磨蹭 duration in different day groups
Mean value
0
5
10
15
20
25
30
01
02
03
04
05
06
07
08
09
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
磨蹭

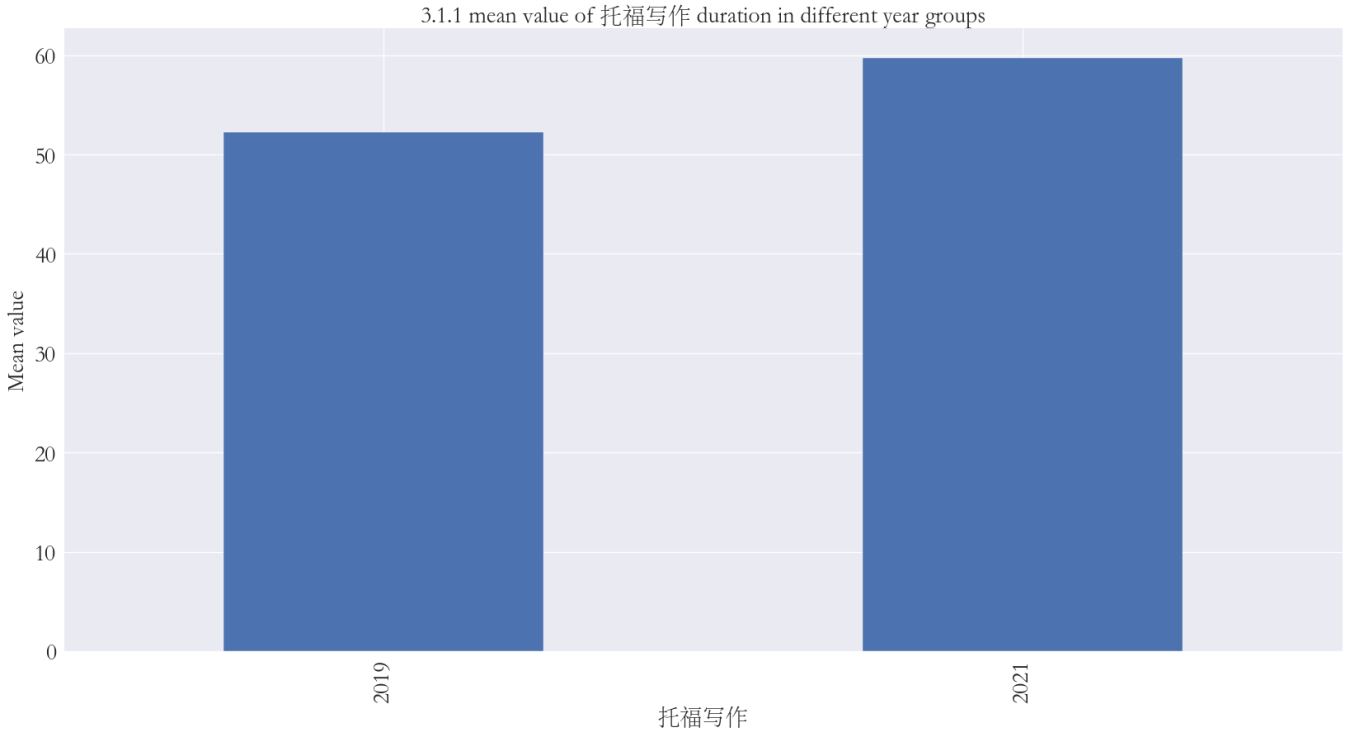

3.1.1 mean value of 磨蹭 duration in different weekday groups
Mean value
0
5
10
15
20
25
30
1
2
3
4
5
6
7
磨蹭

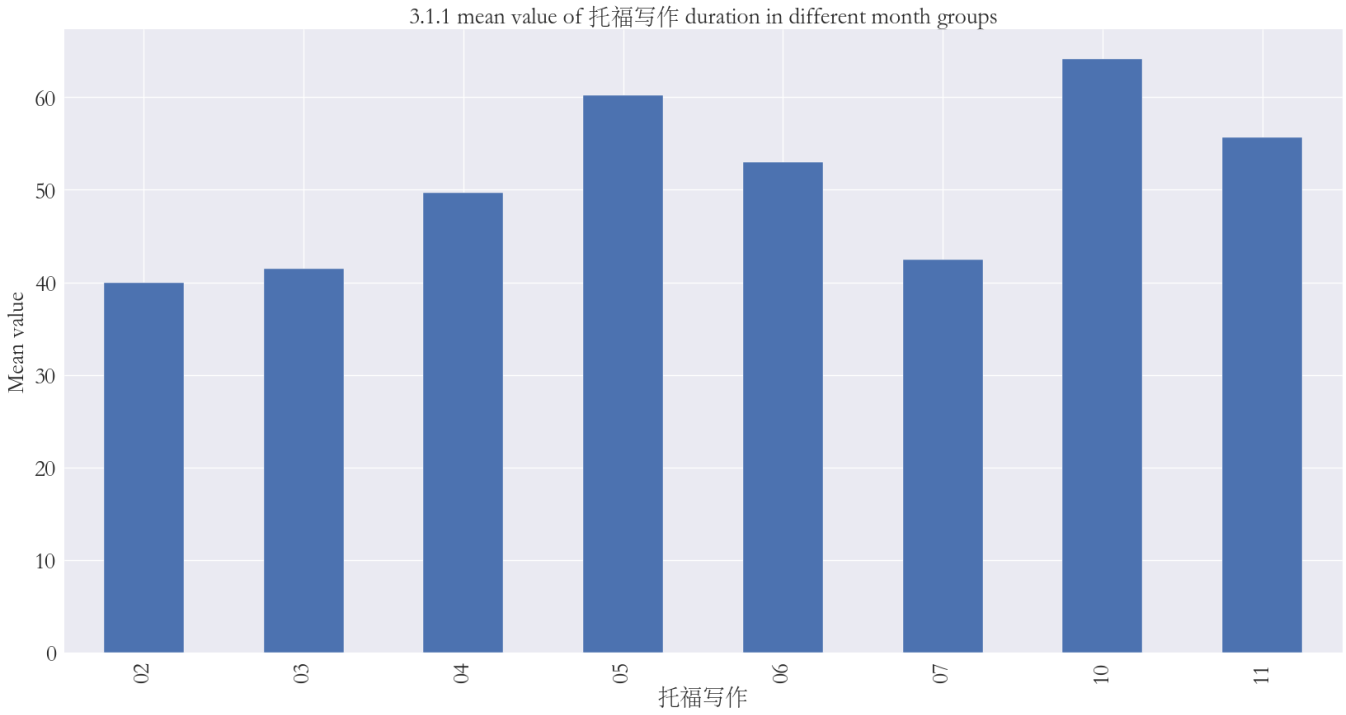

3.1.1 mean value of 晚饭 duration in different day groups
Mean value
0
10
20
30
40
50
01
02
03
04
05
06
07
08
09
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
晚饭

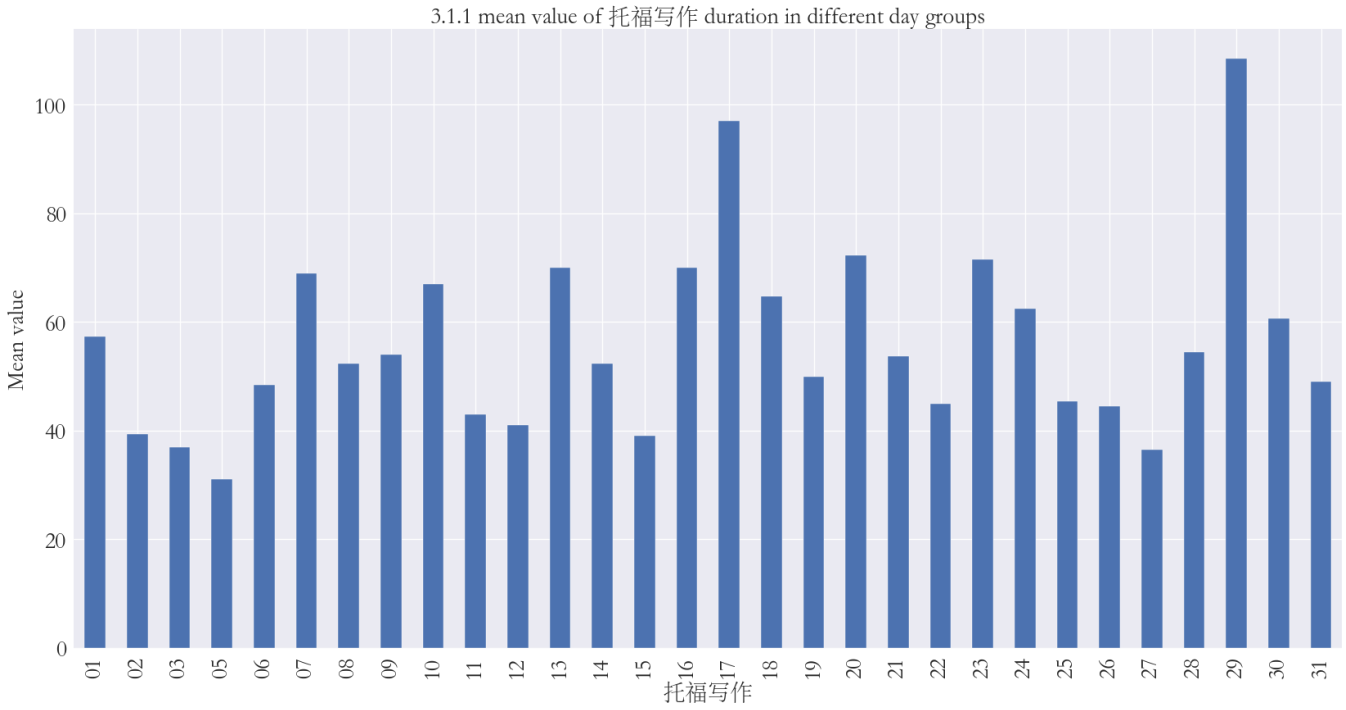

3.1.1 mean value of 晚饭 duration in different weekday groups
Mean value
0
5
10
15
20
25
30
35
40
1
2
3
4
5
6
7
晚饭

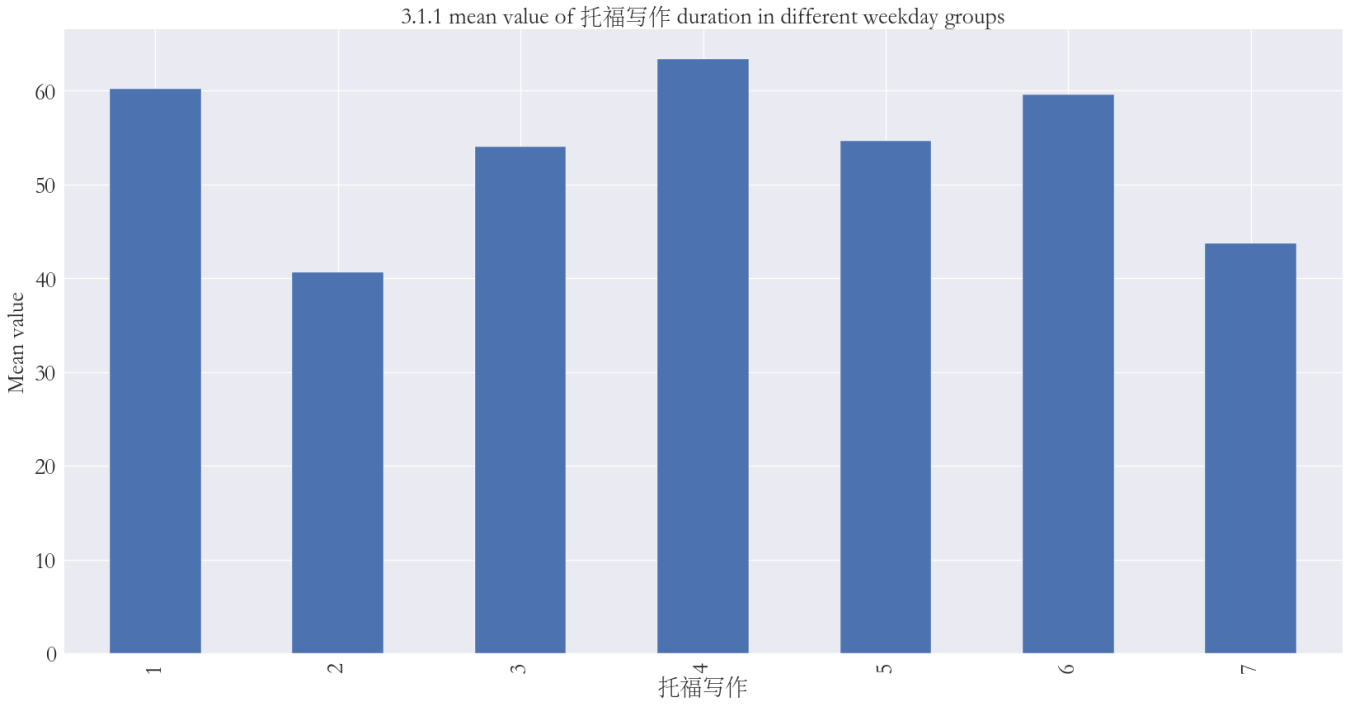

3.1.1 mean value of 晚饭 duration in different mid_hour groups
Mean value
0
10
20
30
40
50
60
70
17
18
19
20
晚饭

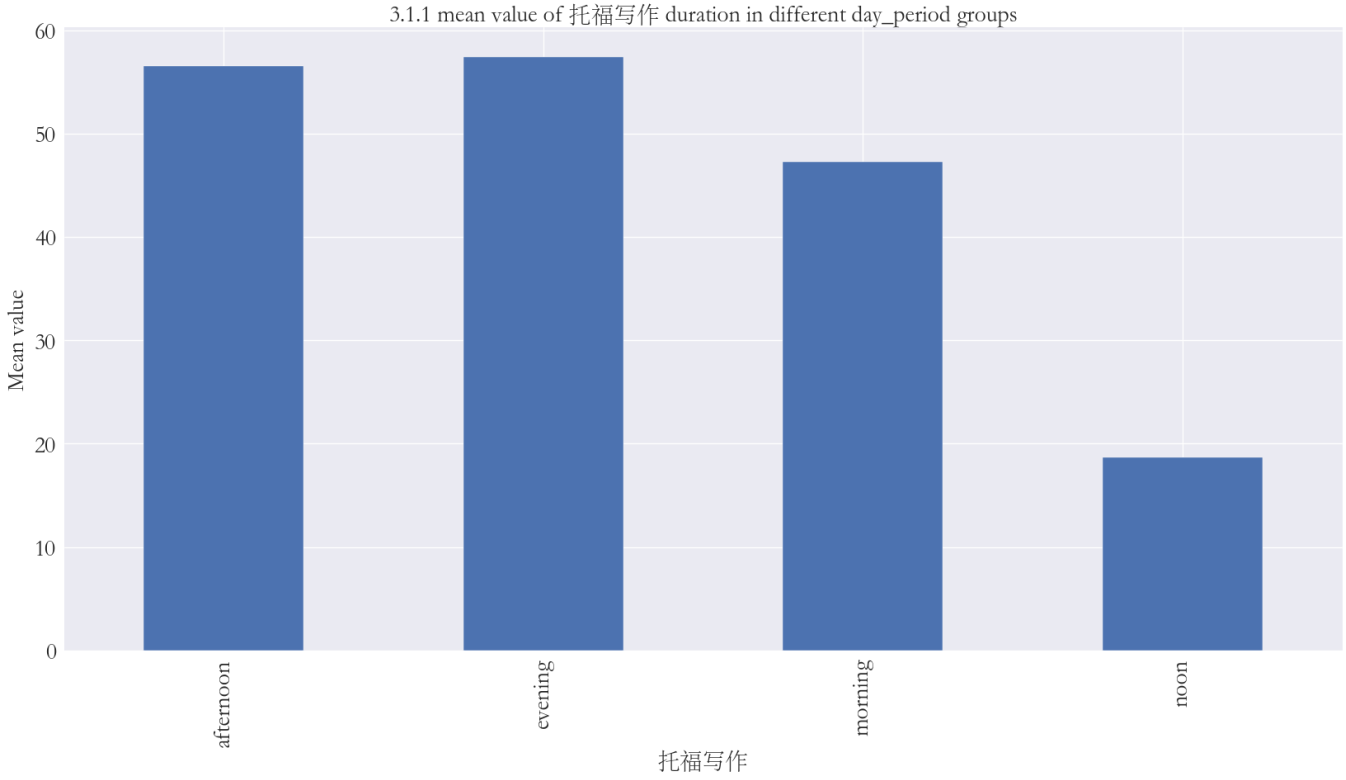

3.1.1 mean value of 晚饭 duration in different day_period groups
Mean value
0
5
10
15
20
25
30
35
40
afternoon
evening
晚饭

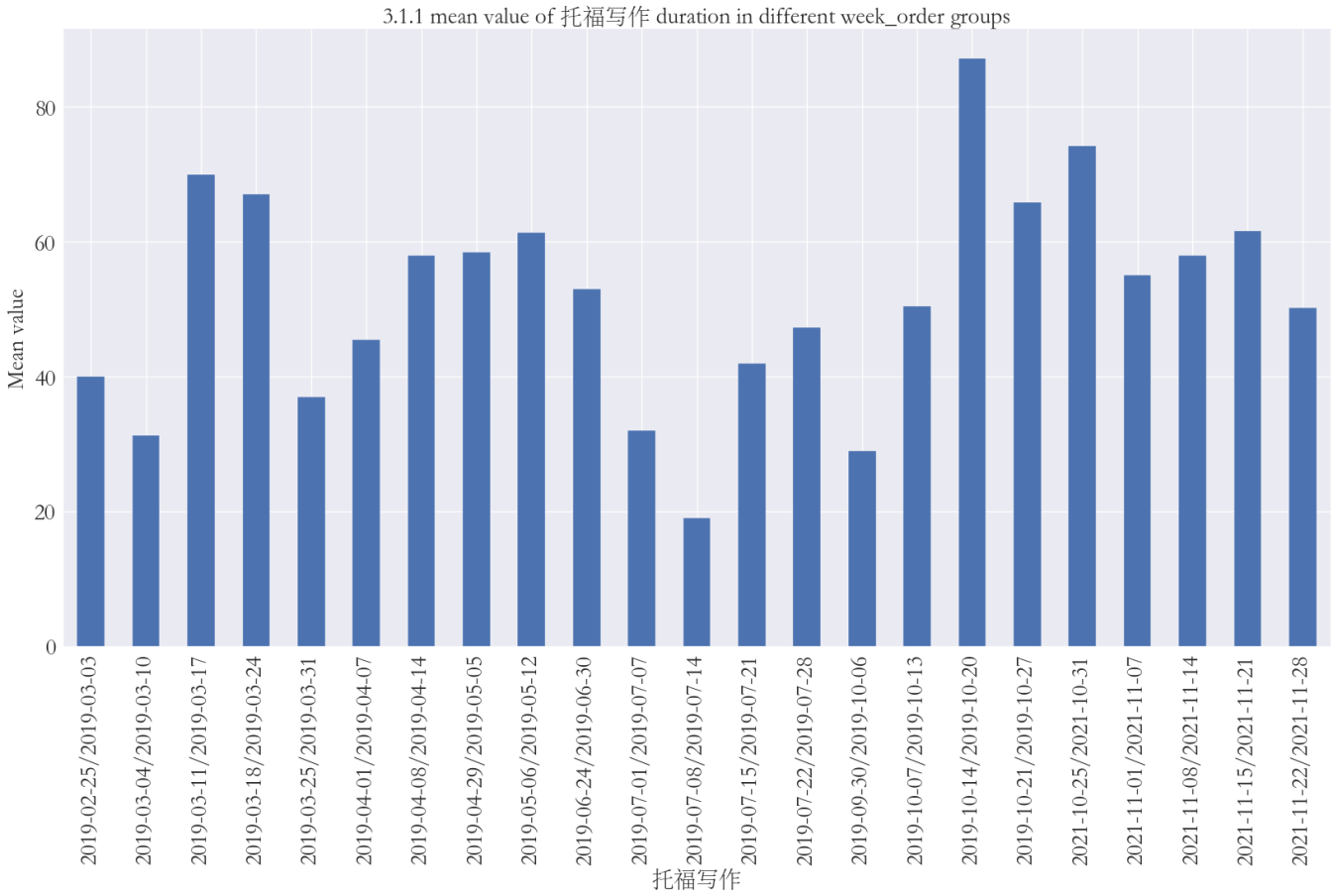

3.1.1 mean value of 托福写作 duration in different year groups
Mean value
0
10
20
30
40
50
60
2019
2021
托福写作

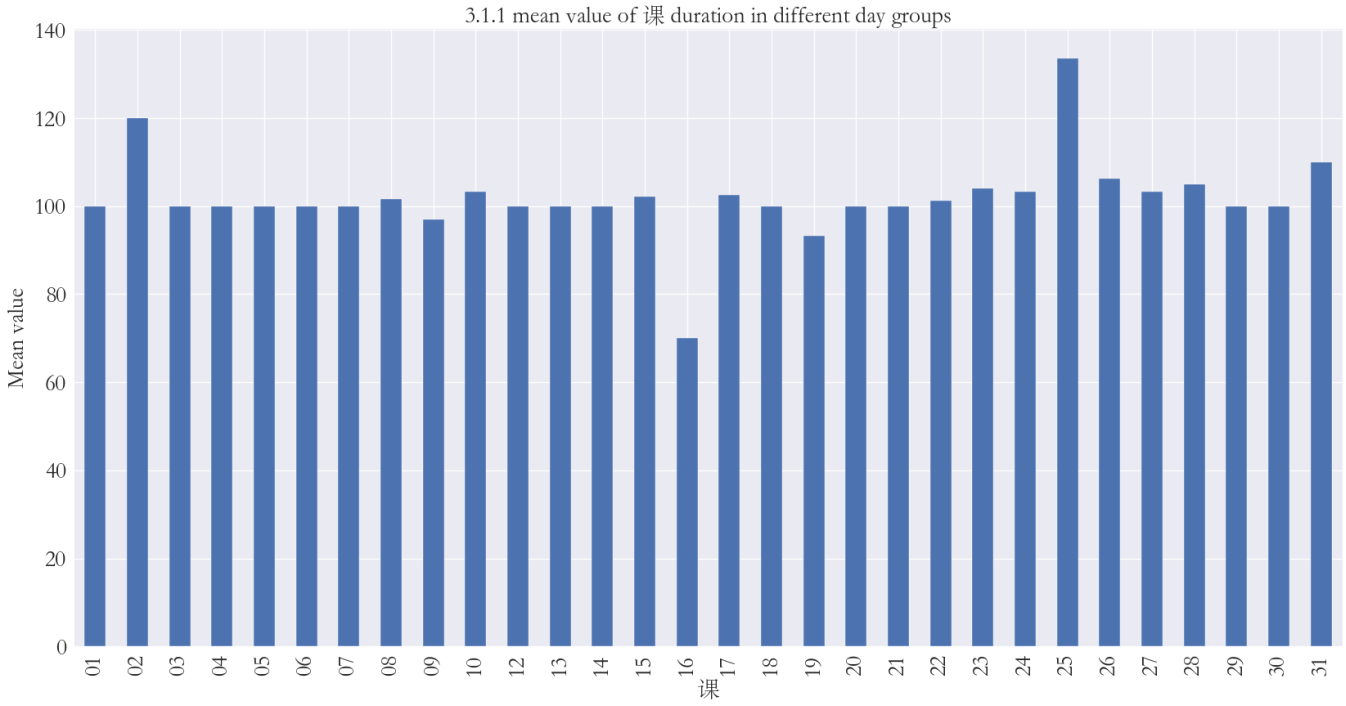

3.1.1 mean value of 托福写作 duration in different month groups
Mean value
0
10
20
30
40
50
60
02
03
04
05
06
07
10
11
托福写作

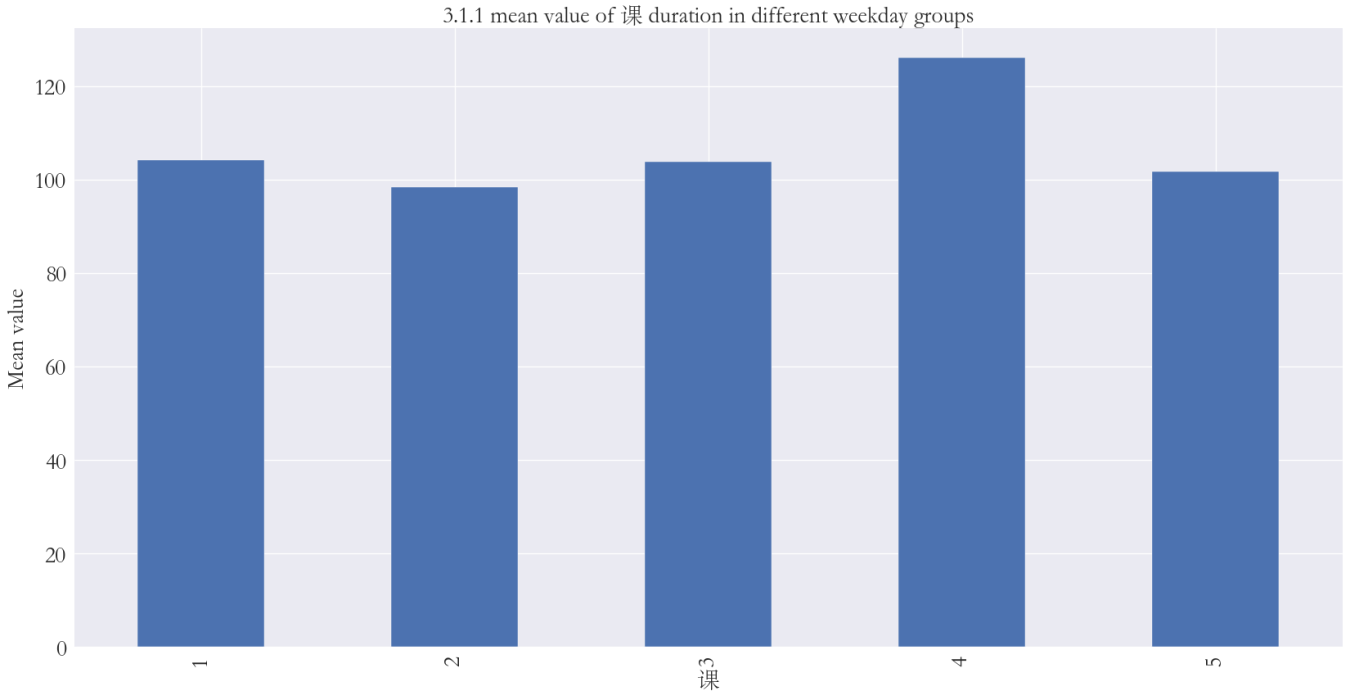

3.1.1 mean value of 托福写作 duration in different day groups
Mean value
0
20
40
60
80
100
01
02
03
05
06
07
08
09
10
11
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
托福写作

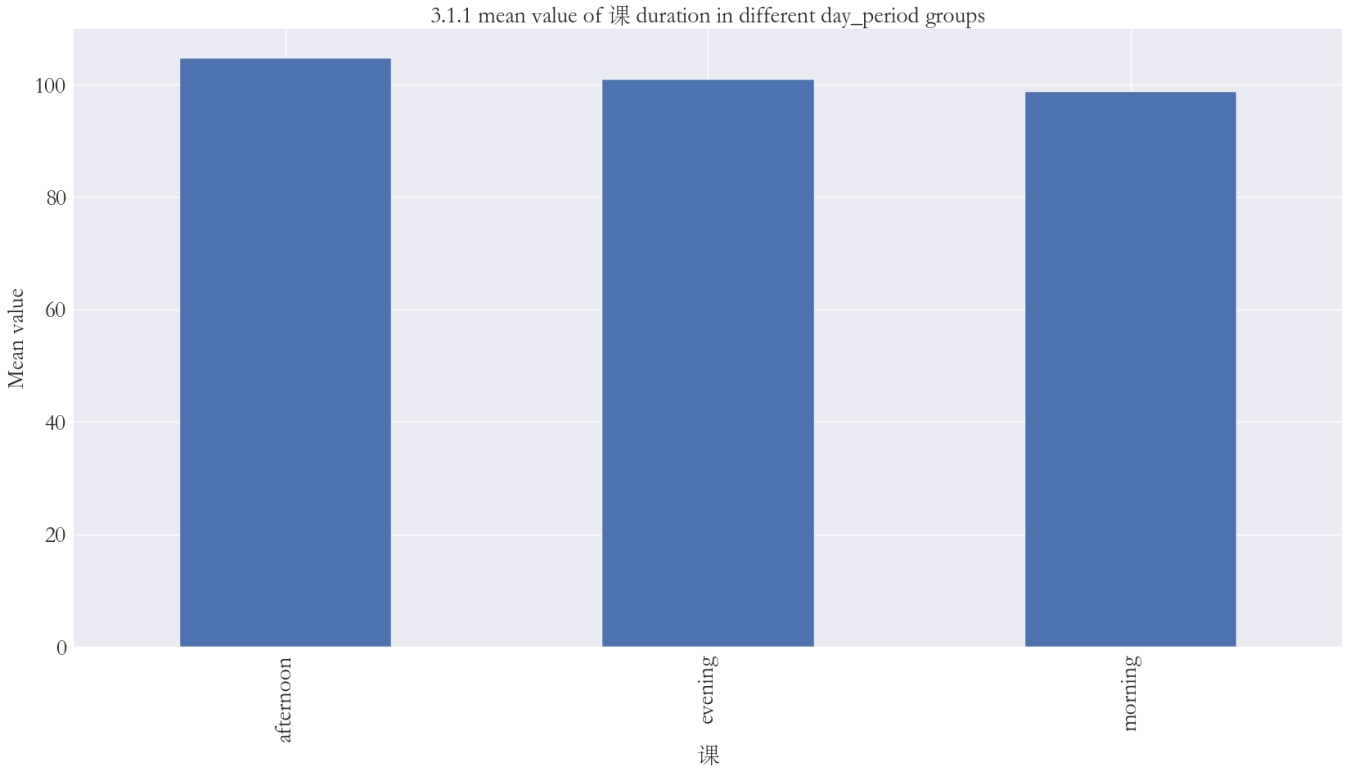

3.1.1 mean value of 托福写作 duration in different weekday groups
Mean value
0
10
20
30
40
50
60
1
2
3
4
5
6
7
托福写作

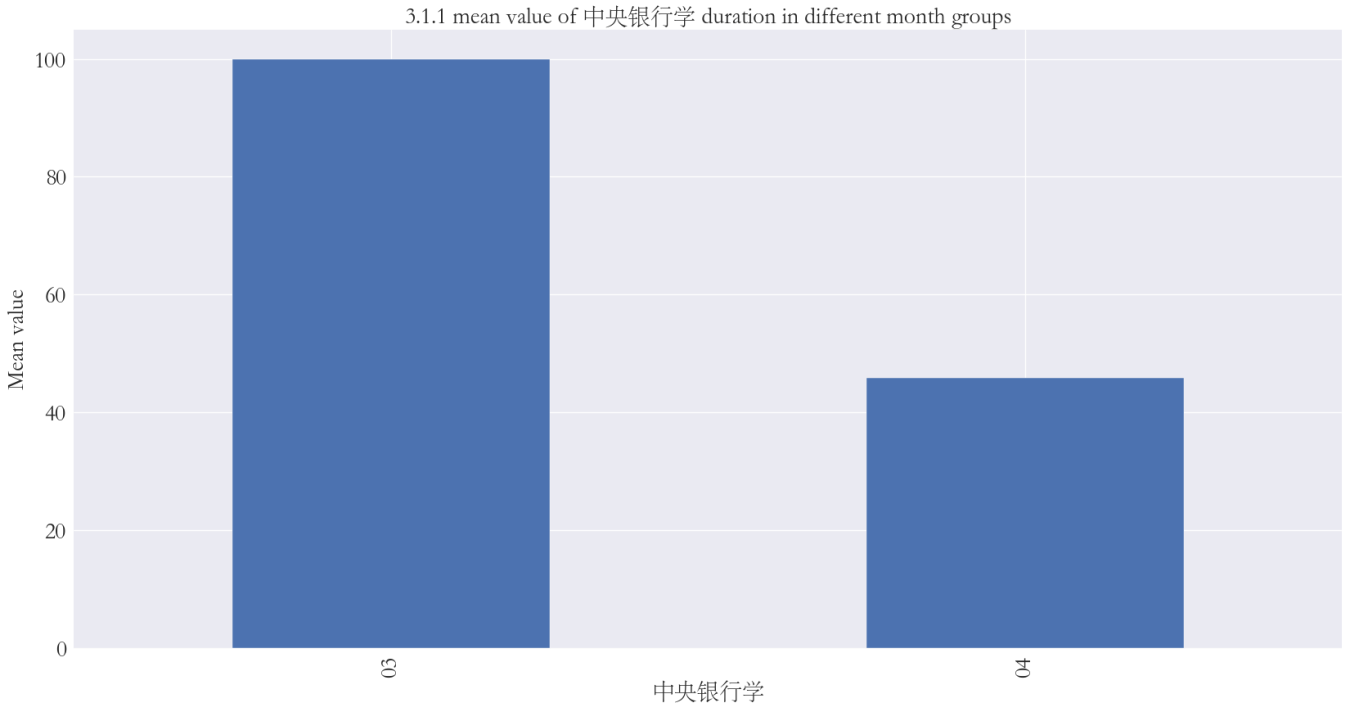

3.1.1 mean value of 托福写作 duration in different day_period groups
Mean value
0
10
20
30
40
50
60
afternoon
evening
morning
noon
托福写作

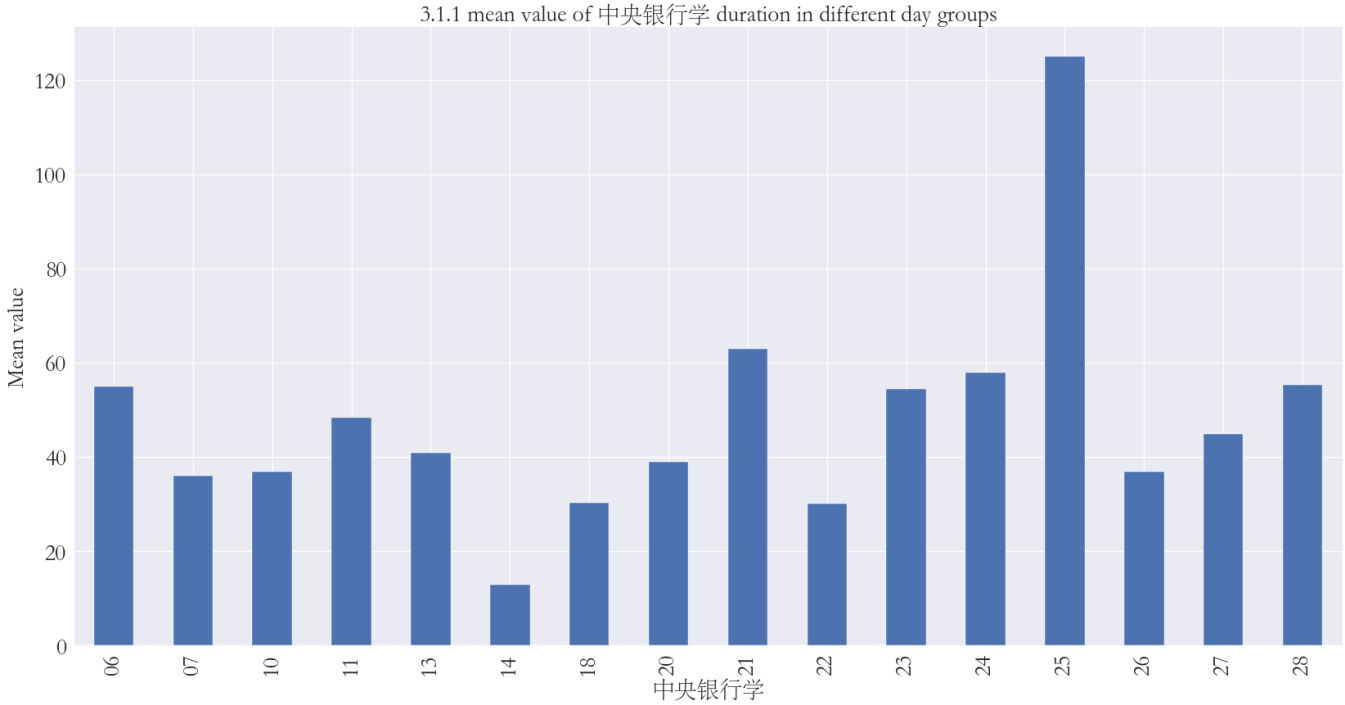

3.1.1 mean value of 托福写作 duration in different week_order groups
Mean value
0
20
40
60
80
2019-02-25/2019-03-03
2019-03-04/2019-03-10
2019-03-11/2019-03-17
2019-03-18/2019-03-24
2019-03-25/2019-03-31
2019-04-01/2019-04-07
2019-04-08/2019-04-14
2019-04-29/2019-05-05
2019-05-06/2019-05-12
2019-06-24/2019-06-30
2019-07-01/2019-07-07
2019-07-08/2019-07-14
2019-07-15/2019-07-21
2019-07-22/2019-07-28
2019-09-30/2019-10-06
2019-10-07/2019-10-13
2019-10-14/2019-10-20
2019-10-21/2019-10-27
2021-10-25/2021-10-31
2021-11-01/2021-11-07
2021-11-08/2021-11-14
2021-11-15/2021-11-21
2021-11-22/2021-11-28
托福写作

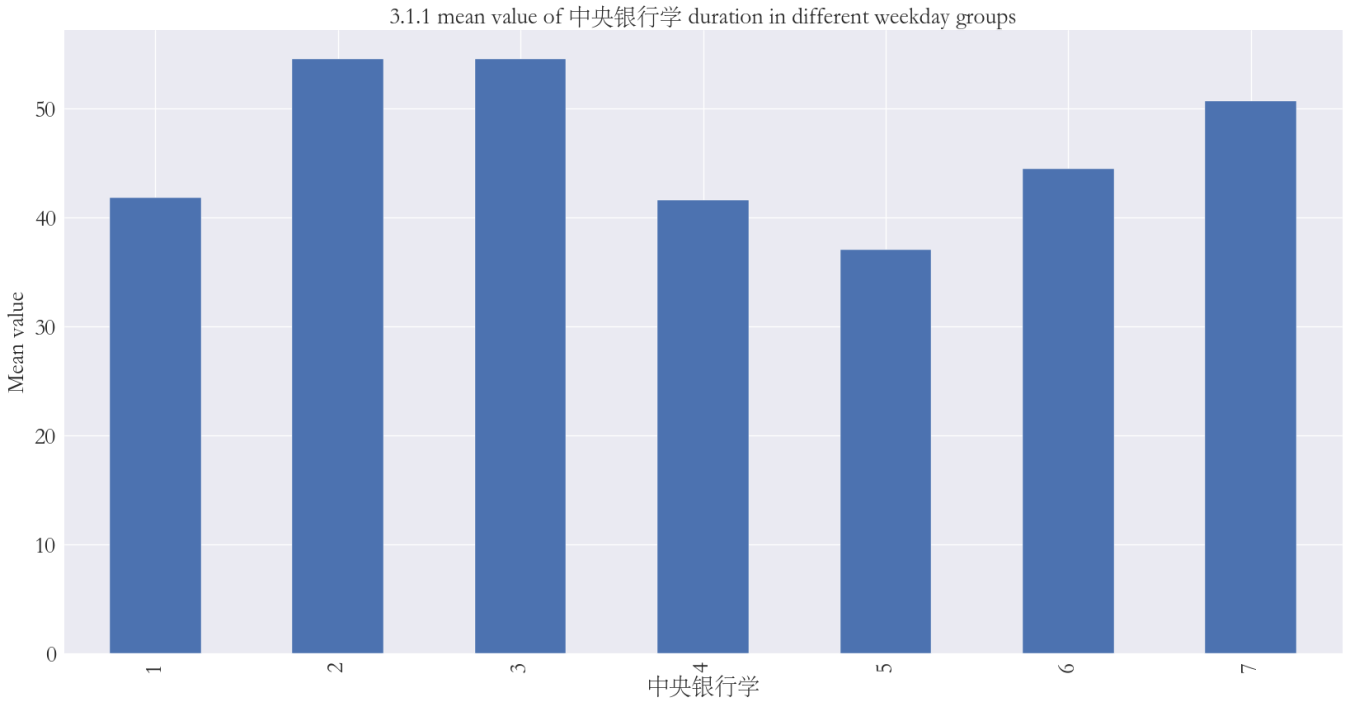

3.1.1 mean value of 课 duration in different day groups
Mean value
0
20
40
60
80
100
120
140
01
02
03
04
05
06
07
08
09
10
12
13
14
15
16
17
18
19
20
21
22
23
24
25
26
27
28
29
30
31
课

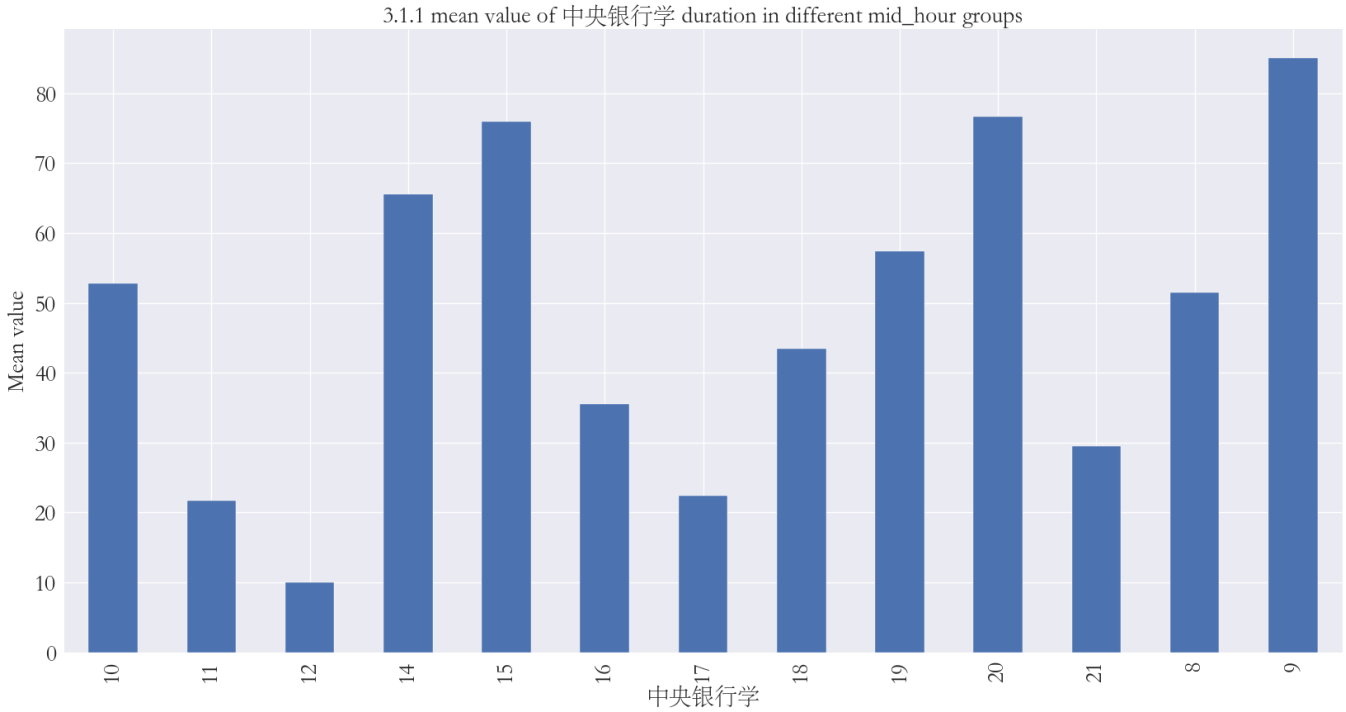

3.1.1 mean value of 课 duration in different weekday groups
Mean value
120
100
80
60
40
20
0
1
2
3
4
5
课

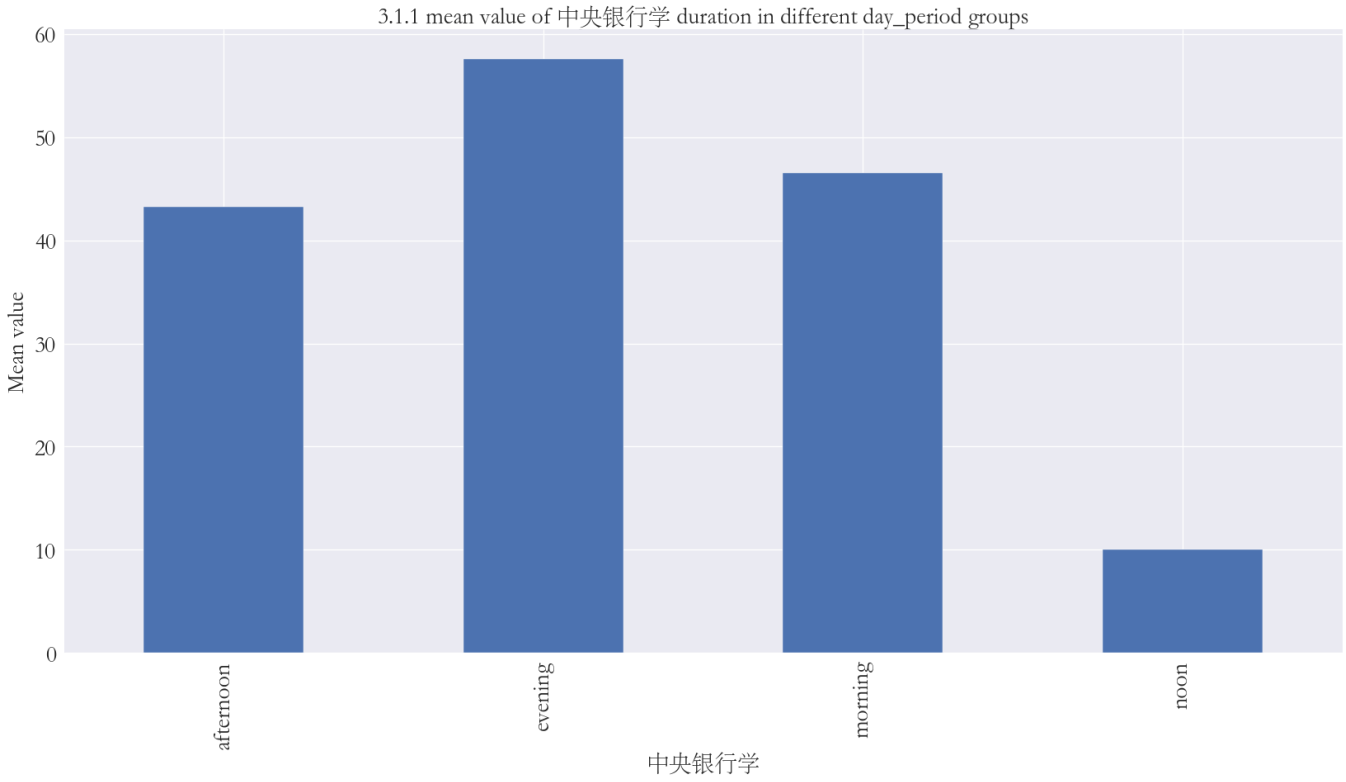

3.1.1 mean value of 课 duration in different day_period groups
Mean value
100
80
60
40
20
0
afternoon
evening
morning
课

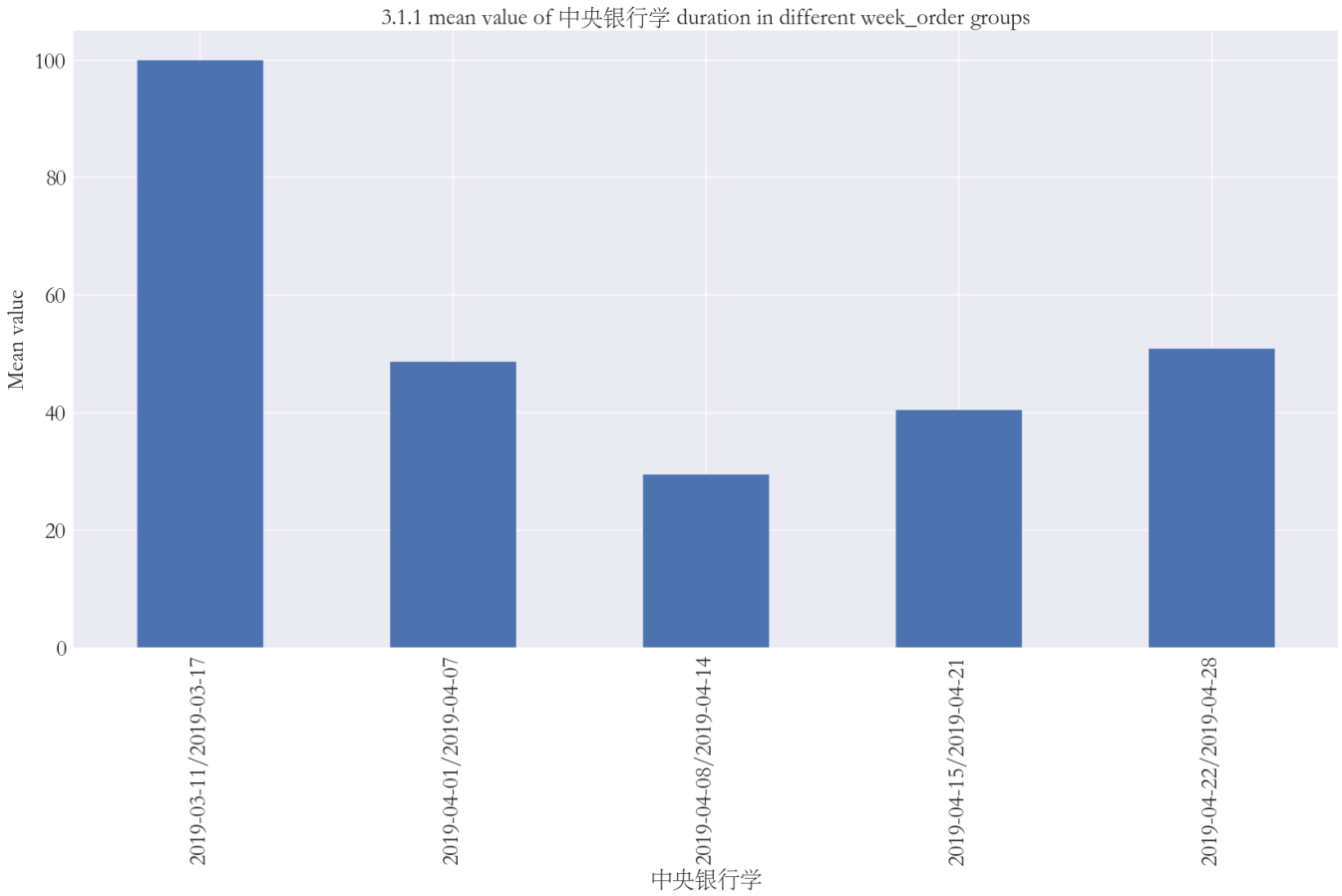

3.1.1 mean value of 中央银行学 duration in different month groups
Mean value
100
80
60
40
20
0
03
04
中央银行学

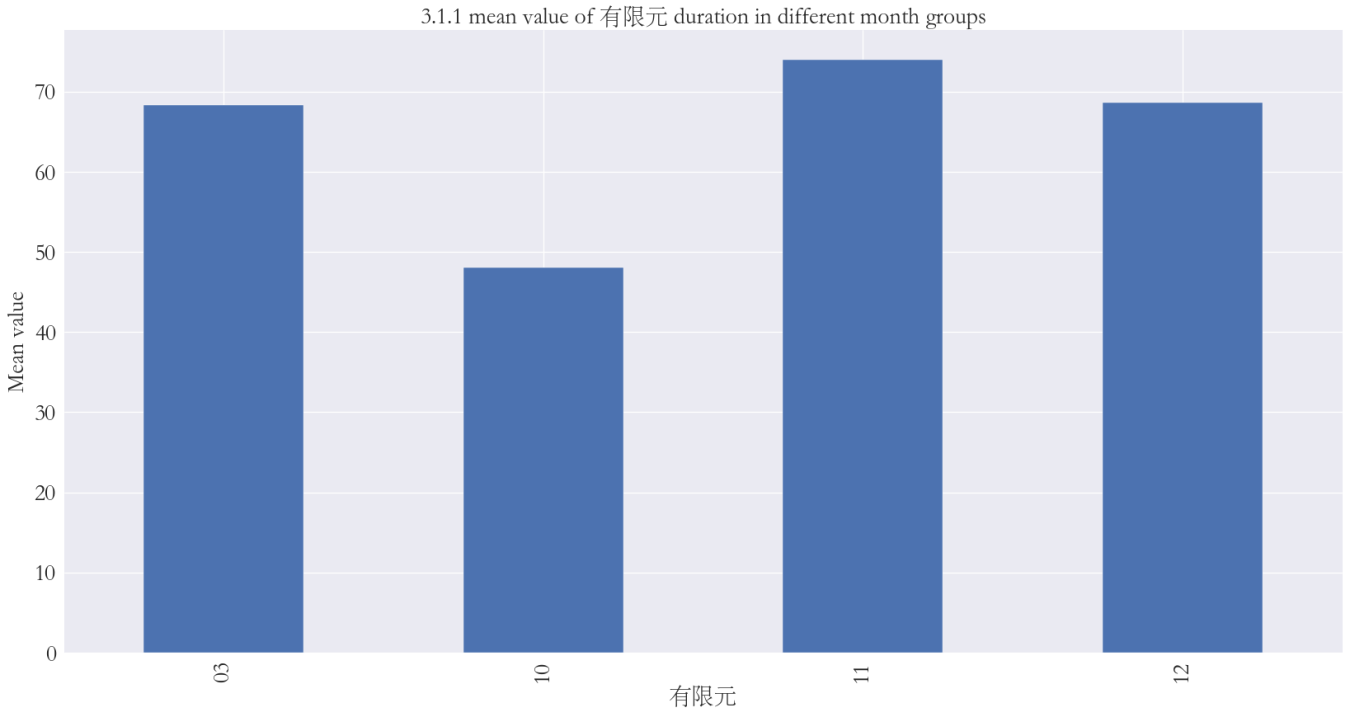

3.1.1 mean value of 中央银行学 duration in different day groups
Mean value
120
100
80
60
40
20
0
06
07
10
11
13
14
18
20
21
22
23
24
25
26
27
28
中央银行学

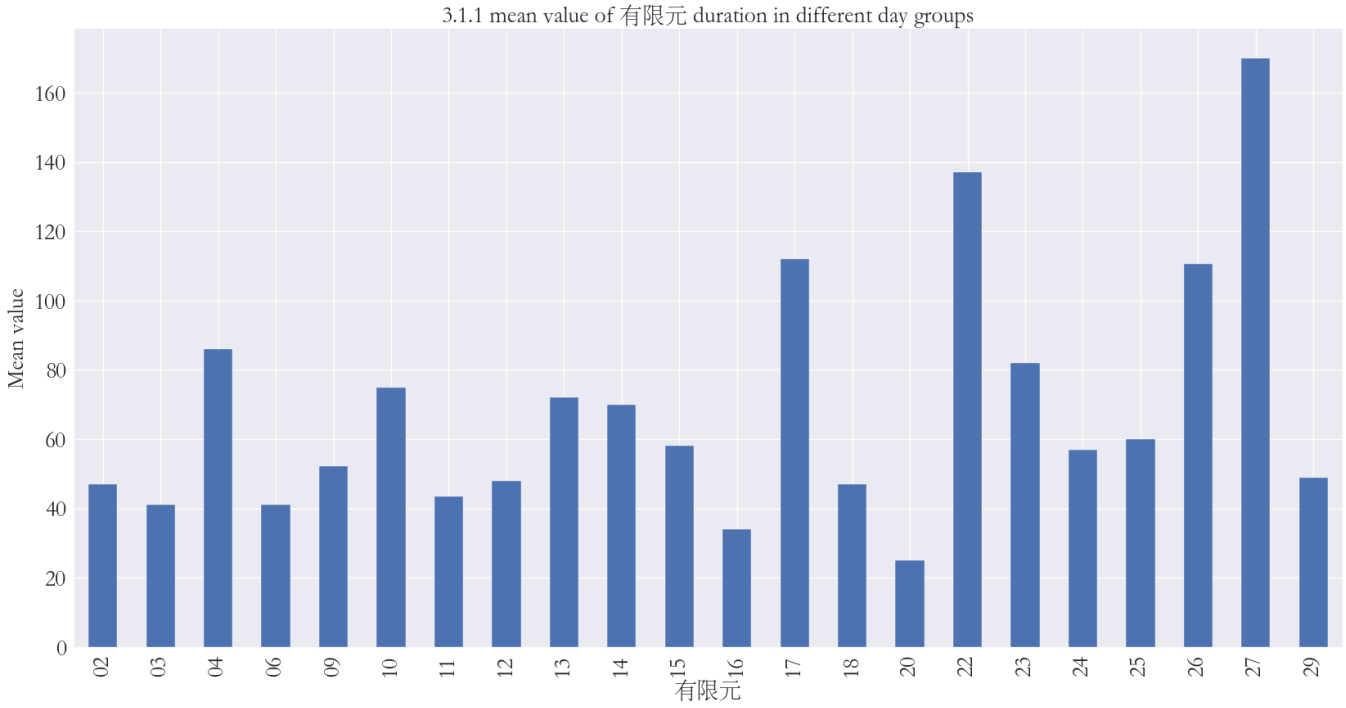

3.1.1 mean value of 中央银行学 duration in different weekday groups
Mean value
0
10
20
30
40
50
1
2
3
4
5
6
7
中央银行学

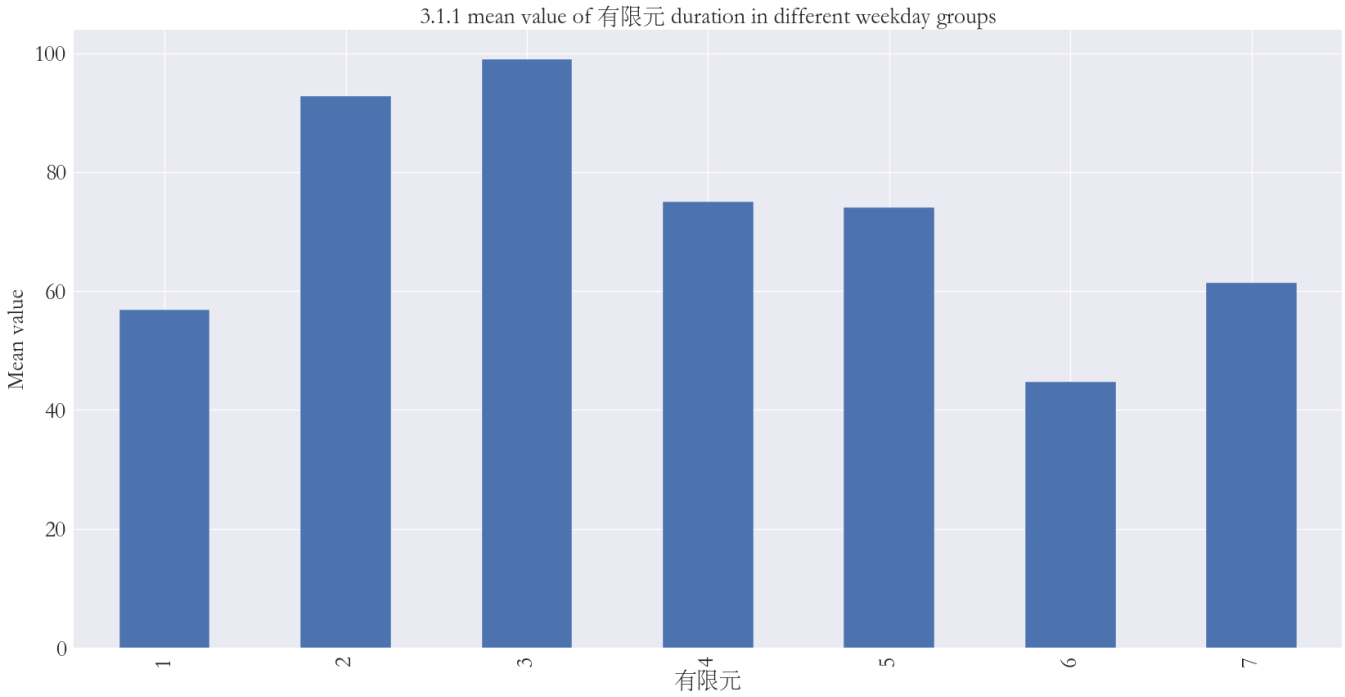

3.1.1 mean value of 中央银行学 duration in different mid_hour groups
Mean value
0
10
20
30
40
50
60
70
80
10
11
12
14
15
16
17
18
19
20
21
8
9
中央银行学

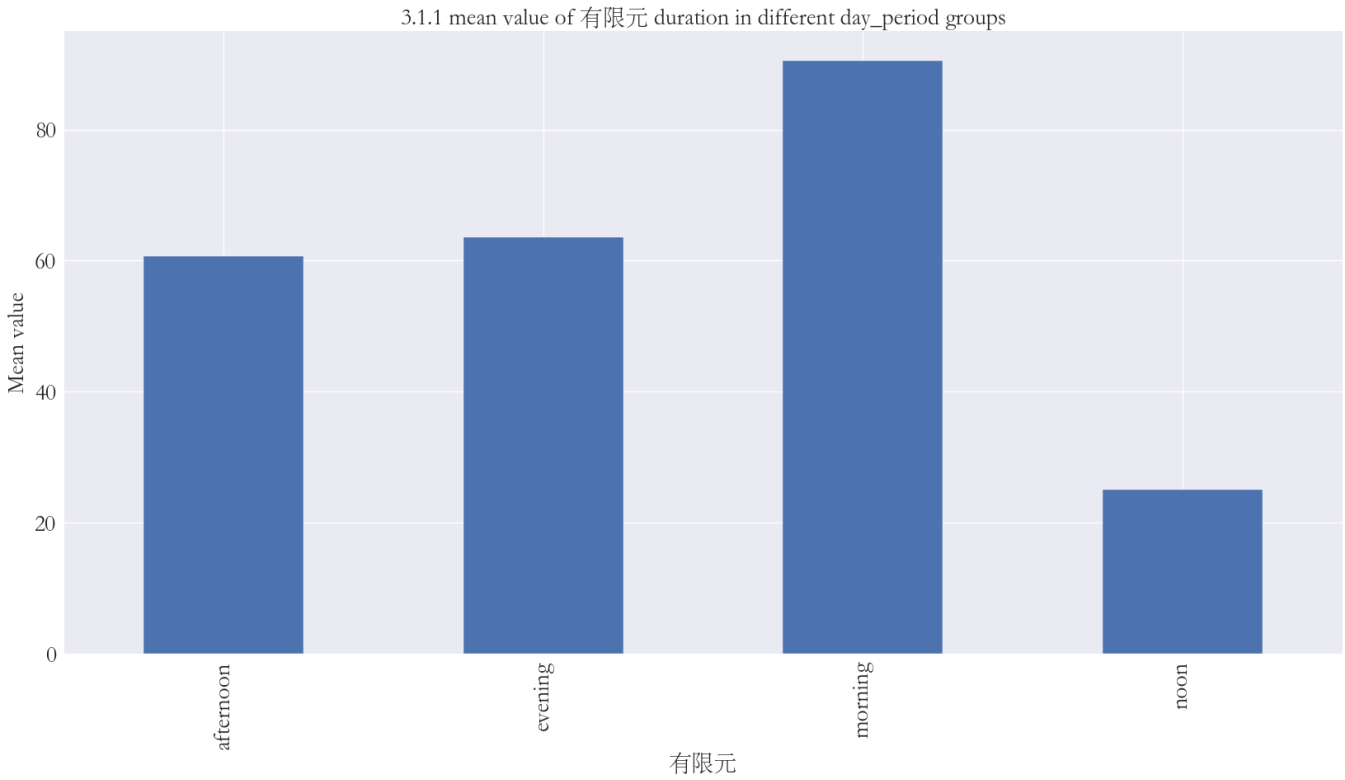

3.1.1 mean value of 中央银行学 duration in different day_period groups
Mean value
60
50
40
30
20
10
0
afternoon
evening
morning
noon
中央银行学


3.1.1 mean value of 中央银行学 duration in different week_order groups
Mean value
100
80
60
40
20
0
2019-03-11/2019-03-17
2019-04-01/2019-04-07
2019-04-08/2019-04-14
2019-04-15/2019-04-21
2019-04-22/2019-04-28
中央银行学


3.1.1 mean value of 有限元 duration in different month groups
Mean value
0
10
20
30
40
50
60
70
03
10
11
12
有限元


3.1.1 mean value of 有限元 duration in different day groups
Mean value
0
20
40
60
80
100
120
140
160
02
03
04
06
09
10
11
12
13
14
15
16
17
18
20
22
23
24
25
26
27
29
有限元


3.1.1 mean value of 有限元 duration in different weekday groups
Mean value
100
80
60
40
20
0
1
2
3
4
5
6
7
有限元


3.1.1 mean value of 有限元 duration in different day_period groups
Mean value
80
60
40
20
0
afternoon
evening
morning
noon
有限元