

In [1]:

```
1 import matplotlib.pyplot as plt
2 import numpy as np
3 import pandas as pd
4 import seaborn as sns
5
6
7 sns.set(font_scale=1.5)
8 %matplotlib inline
```

executed in 888ms, finished 00:12:11 2020-05-21

Формулировка задачи на простом языке:

Выяснить, в какое время дня совершается наибольшее количество поездок.

Формулировка задачи на математическом языке

Исходя из данных о поездках, выяснить, в какое время дня и во сколько именно совершается наибольшее число поездок.

In [2]:

```
1 DATA_DIR = "../data/processed/"
```

executed in 4ms, finished 00:12:11 2020-05-21

In [3]:

```
1 trips = pd.read_csv(DATA_DIR+"trips.csv",
2                       error_bad_lines=False,
3                       index_col=0)
4 trips.head(3)
```

executed in 807ms, finished 00:12:12 2020-05-21

Out[3]:

	trip_id	starttime	stoptime	bikeid	tripduration	from_station_name	to_station
0	431	10/13/2014 10:31	10/13/2014 10:48	SEA00298	985.935	2nd Ave & Spring St	Occide Occident S
1	432	10/13/2014 10:32	10/13/2014 10:48	SEA00195	926.375	2nd Ave & Spring St	Occide Occident S
2	433	10/13/2014 10:33	10/13/2014 10:48	SEA00486	883.831	2nd Ave & Spring St	Occide Occident S

In [4]:

```
1 trips['starttime'] = pd.to_datetime(trips['starttime'])
2 trips['stoptime'] = pd.to_datetime(trips['stoptime'])
```

executed in 38.7s, finished 00:12:53 2020-05-21

In [16]:

```
▶ 1 def get_time_of_day(hour): ↵
   11
   12
▼ 13 def is_weekend(day):
   14     return day > 4
```

executed in 7ms, finished 00:26:16 2020-05-21

In [6]:

```
1 trips['starttime_of_day'] = [get_time_of_day(hour) for hour in trips['starttin
```

executed in 179ms, finished 00:12:54 2020-05-21

Добавим признак, отвечающий за то, является ли день поездки выходным.

In [18]:

```
1 trips['weekend'] = is_weekend(trips.starttime.dt.weekday)
```

executed in 67ms, finished 00:27:53 2020-05-21

In [19]:

```
1 trips.head(3)
```

executed in 61ms, finished 00:27:59 2020-05-21

Out[19]:

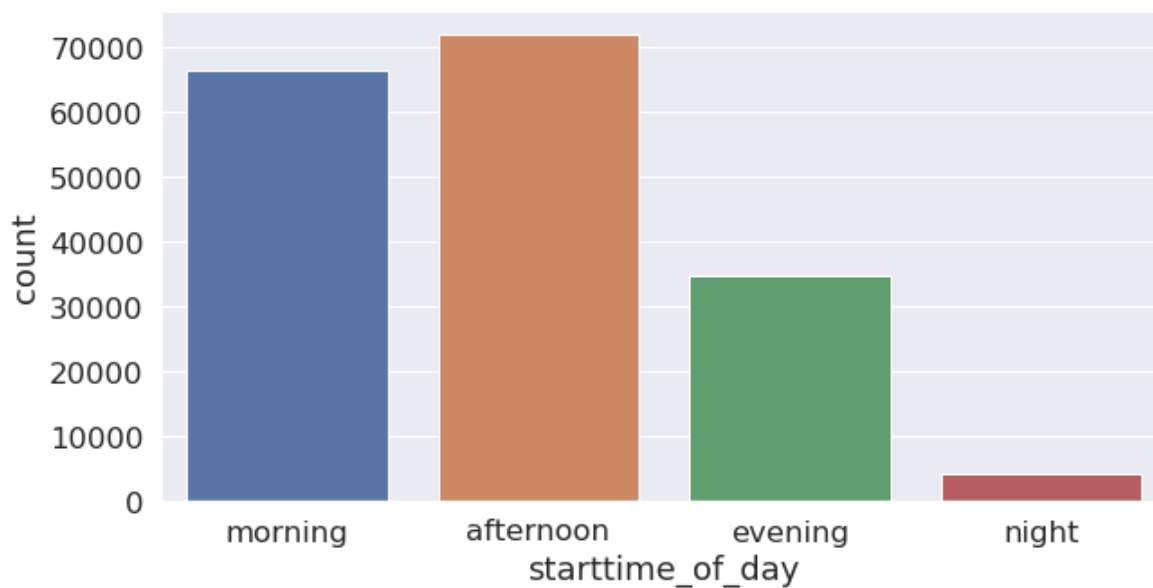
	trip_id	starttime	stoptime	bikeid	tripduration	from_station_name	to_stator
0	431	2014-10-13 10:31:00	2014-10-13 10:48:00	SEA00298	985.935	2nd Ave & Spring St	Occide Occident S
1	432	2014-10-13 10:32:00	2014-10-13 10:48:00	SEA00195	926.375	2nd Ave & Spring St	Occide Occident S
2	433	2014-10-13 10:33:00	2014-10-13 10:48:00	SEA00486	883.831	2nd Ave & Spring St	Occide Occident S

Посмотрим на гистограмму поездок по времени суток за всю историю наблюдений по рабочим дням.

In [20]:

```
1 plt.figure(figsize=(10, 5))
2 sns.countplot(x='starttime_of_day', data=trips[trips.weekend == False])
3 plt.show()
```

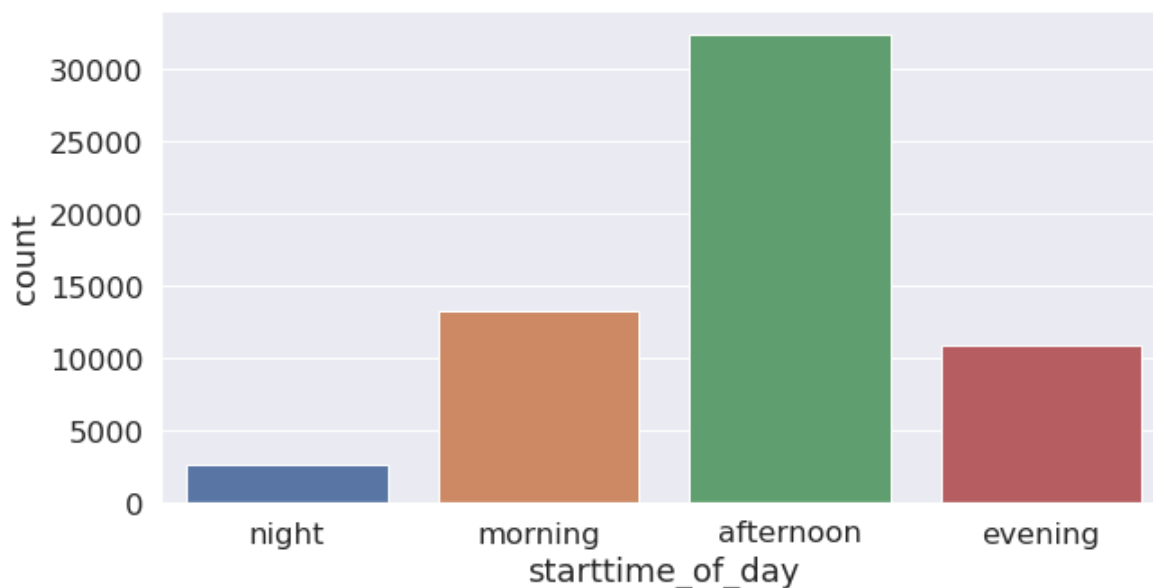
executed in 733ms, finished 00:28:45 2020-05-21



In [21]:

```
1 plt.figure(figsize=(10, 5))
2 sns.countplot(x='starttime_of_day', data=trips[trips.weekend == True])
3 plt.show()
```

executed in 232ms, finished 00:28:59 2020-05-21



Как видим, велосипедами в выходные и рабочие дни пользуются по-разному.

In [22]:

```
1 trips['starttime_of_day_hour'] = [hour for hour in trips['starttime'].dt.hour]
```

executed in 154ms, finished 00:29:33 2020-05-21

In [25]:

```
1 plt.figure(figsize=(10, 5))
2 plt.title('Поездки по часам в рабочие дни')
3 sns.countplot(x='starttime_of_day_hour', data=trips[trips.weekend == False])
4 plt.show()
```

executed in 599ms, finished 00:30:21 2020-05-21



In [26]:

```
1 plt.figure(figsize=(10, 5))
2 plt.title('Поездки по часам в выходные дни')
3 sns.countplot(x='starttime_of_day_hour', data=trips[trips.weekend == True])
4 plt.show()
```

executed in 462ms, finished 00:30:41 2020-05-21



Видим, что пик поездок в рабочие дни приходится на 8 утра и 17 дня - приблизительно время и начало рабочего дня, что вообще говоря объяснимо. В выходные же прокатом активно пользуются днем и вечером до 20 - 21 .