



Capital Bikeshare: Analyse und Prognose der Ausleihvorgänge

20 - Grobe Säuberung der Daten (Trip und Weather)

Hinweis: Die Notebooks sind so aufgebaut, dass sie zu einer Verarbeitungs-Pipeline gehören und in der Reihenfolge der Nummern (Prefixe) ausgeführt sollten, da spätere Notebooks (die mit einer größeren Anfangsnummer) Daten aus den vorherigen Notebooks verwenden. Nur Notebooks mit ganzen 10er-Nummern gehören zur eigentlichen Verarbeitungs-Pipeline.

In [1]:

```
import datetime
import pandas as pd
```

In [2]:

```
# Konstante Werte für Pfade, Dateinamen und andere Vereinbarungen
```

In [3]:

```
DATA_PATH = '../data/'
RAW_TRIPS_FILE = 'trips_raw.pkl'
CLEAN_TRIPS_FILE = 'trips_clean.pkl'
TRIP_COL_LIST_DUPLICATE_CHECK = ['start_ts', 'end_ts', 'start_station_id', 'end_station_id', 'bike_number']
# Hier Änderung: Member Type mit drin lassen:
TRIP_COL_LIST_TO_KEEP = ['start_ts', 'end_ts', 'start_station_id', 'end_station_id', 'bike_number', 'Member type']
RAW_WEATHER_FILE = 'weather_raw.pkl'
ALT_WEATHER_FILE = 'weather_alt.pkl'
CLEAN_WEATHER_FILE = 'weather.pkl'
WEATHER_COLS = ['date', 'hour', 'temperature', 'humidity', 'precipitation', 'windspeed', 'dewpoint', 'pressure']
```

In [4]:

```
# Die (bis auf die Benennung der Merkmale) unveränderten Trip-Daten aus Schritt 1 einlesen
df_trips_raw = pd.read_pickle(DATA_PATH+RAW_TRIPS_FILE)
```

In []:

In [5]:

```
# Identische Einträge der relevanten Merkmale für Entnahme und Rückgabe eines Fahrrades
# identifizieren und entfernen
```

In [6]:

```
def remove_duplicates(df_raw):  
    print('Total trips:', df_raw.shape[0])  
    # Determining duplicates  
    print('Checking for duplicates...')  
    duplicates = df_raw.duplicated(subset=TRIP_COL_LIST_DUPLICATE_CHECK, keep='first')  
    print('Duplicate trips:', duplicates.sum())  
    df_clean = df_raw[~duplicates]  
    print('Total trips remaining:', df_clean.shape[0])  
    print('Done.')  
    return df_clean
```

In []:

In [7]:

```
df_trips = remove_duplicates(df_trips_raw)
```

```
Total trips: 10277677  
Checking for duplicates...  
Duplicate trips: 0  
Total trips remaining: 10277677  
Done.
```

In []:

In [8]:

```
# Überprüfung auf fehlende Werte
```

In [9]:

```
df_trips.isnull().sum()
```

Out[9]:

```
duration          0  
start_ts          0  
end_ts           0  
start_station_id  0  
start_station_name 0  
end_station_id    0  
end_station_name  0  
bike_number       0  
Member type       0  
dtype: int64
```

In []:

In []:

In [10]:

```
# Prüfen ob Rückgabe stets vor der nächsten Entnahme liegt und die nächste Entnahme nach der Rückgabe
```

In [11]:

```
# nach Fahrrädern und Zeitstempeln sortieren
df_trips_sorted = df_trips.sort_values(by=['bike_number', 'start_ts', 'end_ts'])
# Berechne DataFrame mit Spalte, bei der die Entnahmezeit um ein Trip verschoben ist (früher)
df_start_shift = df_trips_sorted.groupby(
    by='bike_number', sort=False)[['start_ts']].shift(-1).rename(
        columns={'start_ts': 'next_start_ts'})
# Hänge DataFrames horizontal zusammen - Sortierreihenfolge darf sich nicht verändert haben
df_check = pd.concat([df_trips_sorted, df_start_shift], axis=1)
```

In [12]:

```
df_check.head()
```

Out[12]:

| | duration | start_ts | end_ts | start_station_id | start_station_name | end_station_id | end |
|---------------|----------|------------------------|------------------------|------------------|----------------------------------|----------------|------|
| 130487 | 14315 | 2015-10-15 10:58:35 | 2015-10-15 14:57:10 | 31219 | 10th St & Constitution Ave NW | 31634 | 3rd |
| 193289 | 1501 | 2016-10-18 10:54:16 | 2016-10-18 11:19:17 | 31292 | 22nd St & Constitution Ave NW | 31292 | (|
| 207012 | 727 | 2016-10-19 12:20:37 | 2016-10-19 12:32:45 | 31618 | 4th & East Capitol St NE | 31618 | 4th |
| 241628 | 1120 | 2016-10-22 12:07:42 | 2016-10-22 12:26:22 | 31249 | Jefferson Memorial | 31249 | Jeff |
| 242374 | 1722 | 2016-10-22 13:01:26 | 2016-10-22 13:30:08 | 31249 | Jefferson Memorial | 31249 | Jeff |

In [13]:

```
# Es gibt einige Trips, bei denen die Rückgabe nach der nächsten Entnahme liegt ... das geht nicht
```

In [14]:

```
df_check.loc[df_check['end_ts']>=df_check['next_start_ts']]
```

Out[14]:

| | duration | start_ts | end_ts | start_station_id | start_station_name | end_station_id | end |
|--------|----------|---------------------|---------------------|------------------|--------------------------------|----------------|---------|
| 290539 | 194 | 2015-11-01 01:07:52 | 2015-11-01 01:11:07 | 31103 | 16th & Harvard St NW | 31107 | |
| 290703 | 599 | 2015-11-01 01:42:03 | 2015-11-01 01:52:02 | 31116 | California St & Florida Ave NW | 31203 | |
| 290517 | 971 | 2015-11-01 01:02:08 | 2015-11-01 01:18:19 | 31280 | 11th & S St NW | 31200 | Mas |
| 290711 | 530 | 2015-11-01 01:43:21 | 2015-11-01 01:52:11 | 31105 | 14th & Harvard St NW | 31503 | Flc |
| 290730 | 371 | 2015-11-01 01:49:02 | 2015-11-01 01:55:14 | 31111 | 10th & U St NW | 31105 | 14 |
| 402149 | 1260 | 2016-11-06 01:11:35 | 2016-11-06 01:32:35 | 31608 | 8th & Eye St SE / Barracks Row | 31623 | Cc |
| 290491 | 1168 | 2015-11-01 00:56:59 | 2015-11-01 01:16:28 | 31221 | 18th & M St NW | 31503 | Flc |
| 402219 | 448 | 2016-11-06 01:43:18 | 2016-11-06 01:50:47 | 31268 | 12th & U St NW | 31102 | 1' |
| 290504 | 353 | 2015-11-01 01:00:28 | 2015-11-01 01:06:21 | 31509 | New Jersey Ave & R St NW | 31202 | |
| 290505 | 1341 | 2015-11-01 01:00:29 | 2015-11-01 01:22:51 | 31237 | 25th St & Pennsylvania Ave NW | 31237 | Pe |
| 290595 | 386 | 2015-11-01 01:15:51 | 2015-11-01 01:22:17 | 31102 | 11th & Kenyon St NW | 31109 | |
| 290567 | 442 | 2015-11-01 01:12:30 | 2015-11-01 01:19:53 | 31104 | Adams Mill & Columbia Rd NW | 31267 | Mas |
| 290606 | 732 | 2015-11-01 01:18:29 | 2015-11-01 01:30:41 | 31281 | 8th & O St NW | 31104 | C |
| 402154 | 723 | 2016-11-06 01:13:47 | 2016-11-06 01:25:51 | 31400 | Georgia & New Hampshire Ave NW | 31229 | I |
| 290649 | 961 | 2015-11-01 01:28:08 | 2015-11-01 01:44:09 | 31404 | 9th & Upshur St NW | 31114 | 18th |
| 290534 | 2604 | 2015-11-01 01:06:44 | 2015-11-01 01:50:09 | 31278 | 18th & R St NW | 31229 | I |
| 290565 | 1565 | 2015-11-01 01:11:27 | 2015-11-01 01:37:32 | 31114 | 18th St & Wyoming Ave NW | 31636 | Ne N |

| | duration | start_ts | end_ts | start_station_id | start_station_name | end_station_id | end_station_name |
|--------|----------|---------------------|---------------------|------------------|-------------------------------|----------------|------------------|
| 290499 | 1349 | 2015-11-01 00:59:13 | 2015-11-01 01:21:42 | 31201 | 15th & P St NW | 31606 | Pe... |
| 290545 | 660 | 2015-11-01 01:08:35 | 2015-11-01 01:19:36 | 31212 | 21st & M St NW | 31224 | |
| 290589 | 788 | 2015-11-01 01:15:24 | 2015-11-01 01:28:32 | 31280 | 11th & S St NW | 31200 | Mas... |
| 290688 | 524 | 2015-11-01 01:37:31 | 2015-11-01 01:46:15 | 31102 | 11th & Kenyon St NW | 31268 | |
| 648030 | 770 | 2017-08-20 11:24:18 | 2017-08-20 11:37:08 | 31321 | 15th St & Constitution Ave NW | 31321 | (... |

In [15]:

```
# Rückgabe-Zeitpunkt kurz vor den nächsten Entnahmezeitpunkt setzen
df_check.loc[df_check['end_ts']>=df_check['next_start_ts'], 'end_ts'] = \
    df_check.loc[df_check['end_ts']>=df_check['next_start_ts']]['next_start_ts'] - pd.T
imedelta(1, 'second')
```

In [16]:

```
# Überprüfen
df_check.loc[df_check['end_ts']>=df_check['next_start_ts']]
```

Out[16]:

| duration | start_ts | end_ts | start_station_id | start_station_name | end_station_id | end_station_name |
|----------|----------|--------|------------------|--------------------|----------------|------------------|
|----------|----------|--------|------------------|--------------------|----------------|------------------|

In []:

In [17]:

```
# Prüfen, ob Rückgabezeit stets nach Entnahmezeit  
df_check[df_check['start_ts'] >= df_check['end_ts']]
```

Out[17]:

| | duration | start_ts | end_ts | start_station_id | start_station_name | end_station_id | end |
|---------------|----------|------------------------|------------------------|------------------|-----------------------------------|----------------|---------|
| 290777 | 1677 | 2015-11-01 01:59:13 | 2015-11-01 01:27:11 | 31401 | 14th St & Spring Rd NW | 31611 | |
| 290773 | 718 | 2015-11-01 01:58:26 | 2015-11-01 01:10:25 | 31245 | 7th & R St NW / Shaw Library | 31603 | |
| 290729 | 1067 | 2015-11-01 01:49:02 | 2015-11-01 01:06:50 | 31247 | Jefferson Dr & 14th St SW | 31634 | 3rd |
| 290743 | 1911 | 2015-11-01 01:50:33 | 2015-11-01 01:22:25 | 31102 | 11th & Kenyon St NW | 31620 | |
| 402249 | 1175 | 2016-11-06 01:55:59 | 2016-11-06 01:15:34 | 31102 | 11th & Kenyon St NW | 31614 | |
| 402240 | 770 | 2016-11-06 01:51:34 | 2016-11-06 01:04:24 | 31254 | 15th & K St NW | 31202 | |
| 290768 | 338 | 2015-11-01 01:56:35 | 2015-11-01 01:02:13 | 31116 | California St & Florida Ave NW | 31119 | 14 |
| 431851 | 806 | 2017-11-05 01:53:32 | 2017-11-05 01:06:59 | 31266 | 11th & M St NW | 31126 | |
| 290728 | 1078 | 2015-11-01 01:48:59 | 2015-11-01 01:06:58 | 31247 | Jefferson Dr & 14th St SW | 31634 | 3rd |
| 290742 | 1032 | 2015-11-01 01:50:28 | 2015-11-01 01:07:41 | 31600 | 5th & K St NW | 31622 | |
| 431839 | 2632 | 2017-11-05 01:47:44 | 2017-11-05 01:31:37 | 31275 | New Hampshire Ave & 24th St NW | 31275 | I Av |
| 290763 | 512 | 2015-11-01 01:54:56 | 2015-11-01 01:03:28 | 31110 | 20th St & Florida Ave NW | 31109 | |
| 290778 | 170 | 2015-11-01 01:59:19 | 2015-11-01 01:02:09 | 31611 | 13th & H St NE | 31512 | Ne |
| 431854 | 703 | 2017-11-05 01:54:33 | 2017-11-05 01:06:17 | 31401 | 14th St & Spring Rd NW | 31101 | |
| 290741 | 767 | 2015-11-01 01:50:21 | 2015-11-01 01:03:09 | 31228 | 8th & H St NW | 31108 | |
| 402188 | 3578 | 2016-11-06 01:28:02 | 2016-11-06 01:27:40 | 31314 | 34th & Water St NW | 31314 | |
| 290725 | 1912 | 2015-11-01 01:47:58 | 2015-11-01 01:19:50 | 31106 | Calvert & Biltmore St NW | 31106 | Ce |

| | duration | start_ts | end_ts | start_station_id | start_station_name | end_station_id | end |
|--------|----------|------------------------|------------------------|------------------|--|----------------|---------|
| 290760 | 1485 | 2015-11-01 01:54:02 | 2015-11-01 01:18:47 | 31121 | Calvert St & Woodley PI NW | 31108 | |
| 290504 | 353 | 2015-11-01 01:00:28 | 2015-11-01 01:00:28 | 31509 | New Jersey Ave & R St NW | 31202 | |
| 402187 | 3583 | 2016-11-06 01:27:57 | 2016-11-06 01:27:40 | 31314 | 34th & Water St NW | 31314 | |
| 402237 | 2685 | 2016-11-06 01:50:47 | 2016-11-06 01:35:33 | 31258 | Lincoln Memorial | 31286 | |
| 402236 | 2734 | 2016-11-06 01:49:42 | 2016-11-06 01:35:16 | 31258 | Lincoln Memorial | 31286 | |
| 290771 | 1088 | 2015-11-01 01:58:13 | 2015-11-01 01:16:21 | 31245 | 7th & R St NW / Shaw Library | 31612 | L |
| 290747 | 1563 | 2015-11-01 01:51:37 | 2015-11-01 01:17:41 | 31237 | 25th St & Pennsylvania Ave NW | 31237 | Pe |
| 431855 | 254 | 2017-11-05 01:56:50 | 2017-11-05 01:01:04 | 31615 | 6th & H St NE | 31627 | |
| 290776 | 187 | 2015-11-01 01:58:56 | 2015-11-01 01:02:03 | 31611 | 13th & H St NE | 31512 | Ne |
| 402209 | 2978 | 2016-11-06 01:36:13 | 2016-11-06 01:25:51 | 31277 | 17th & G St NW | 31289 | He L |
| 402198 | 3472 | 2016-11-06 01:31:51 | 2016-11-06 01:29:44 | 31277 | 17th & G St NW | 31292 | (|
| 402255 | 474 | 2016-11-06 01:57:47 | 2016-11-06 01:05:41 | 31509 | New Jersey Ave & R St NW | 31600 | |
| 290735 | 783 | 2015-11-01 01:49:33 | 2015-11-01 01:02:36 | 31241 | Thomas Circle | 31254 | |
| 290770 | 1389 | 2015-11-01 01:57:32 | 2015-11-01 01:20:41 | 31200 | Massachusetts Ave & Dupont Circle NW | 31226 | Wis |
| 431821 | 1348 | 2017-11-05 01:39:33 | 2017-11-05 01:02:02 | 31015 | Rosslyn Metro / Wilson Blvd & Ft Myer Dr | 31267 | Mas |
| 402258 | 333 | 2016-11-06 01:59:53 | 2016-11-06 01:05:26 | 31114 | 18th St & Wyoming Ave NW | 31111 | |
| 402241 | 2631 | 2016-11-06 01:51:34 | 2016-11-06 01:35:26 | 31258 | Lincoln Memorial | 31286 | |
| 290746 | 936 | 2015-11-01 01:51:23 | 2015-11-01 01:07:00 | 31203 | 14th & Rhode Island Ave NW | 31638 | |

| | duration | start_ts | end_ts | start_station_id | start_station_name | end_station_id | end |
|--------|----------|------------------------|------------------------|------------------|---|----------------|---------|
| 402254 | 426 | 2016-11-06 01:57:42 | 2016-11-06 01:04:48 | 31245 | 7th & R St NW / Shaw Library | 31505 | Ecki |
| 290744 | 1848 | 2015-11-01 01:50:42 | 2015-11-01 01:21:31 | 31603 | 1st & M St NE | 31623 | Co |
| 290767 | 355 | 2015-11-01 01:56:07 | 2015-11-01 01:02:03 | 31116 | California St & Florida Ave NW | 31119 | 14 |
| 402194 | 3312 | 2016-11-06 01:30:30 | 2016-11-06 01:25:42 | 31277 | 17th & G St NW | 31289 | He L |
| 290707 | 1721 | 2015-11-01 01:42:38 | 2015-11-01 01:11:19 | 31228 | 8th & H St NW | 31006 | : |
| 290765 | 507 | 2015-11-01 01:55:19 | 2015-11-01 01:03:46 | 31109 | 7th & T St NW | 31505 | Ecki |
| 290769 | 327 | 2015-11-01 01:57:26 | 2015-11-01 01:02:53 | 31023 | Fairfax Dr & Wilson Blvd | 31035 | |
| 290749 | 609 | 2015-11-01 01:51:47 | 2015-11-01 01:01:57 | 31268 | 12th & U St NW | 31203 | |
| 290758 | 1397 | 2015-11-01 01:53:42 | 2015-11-01 01:16:59 | 31200 | Massachusetts Ave & Dupont Circle NW | 31244 | |
| 402256 | 1523 | 2016-11-06 01:58:40 | 2016-11-06 01:24:03 | 31101 | 14th & V St NW | 31249 | Jeff |
| 402246 | 637 | 2016-11-06 01:53:43 | 2016-11-06 01:04:20 | 31231 | 14th & D St NW / Ronald Reagan Building | 31108 | |
| 402248 | 766 | 2016-11-06 01:55:43 | 2016-11-06 01:08:30 | 31223 | Convention Center / 7th & M St NW | 31640 | Ma |
| 431857 | 1044 | 2017-11-05 01:59:09 | 2017-11-05 01:16:34 | 31609 | Maine Ave & 7th St SW | 31109 | |
| 431856 | 1081 | 2017-11-05 01:58:02 | 2017-11-05 01:16:03 | 31111 | 10th & U St NW | 31108 | |
| 402253 | 533 | 2016-11-06 01:56:56 | 2016-11-06 01:05:49 | 31121 | Calvert St & Woodley PI NW | 31107 | |
| 402250 | 525 | 2016-11-06 01:56:10 | 2016-11-06 01:04:55 | 31600 | 5th & K St NW | 31245 | . |
| 402202 | 3348 | 2016-11-06 01:33:30 | 2016-11-06 01:29:18 | 31277 | 17th & G St NW | 31292 | (|
| 431843 | 2499 | 2017-11-05 01:49:58 | 2017-11-05 01:31:37 | 31275 | New Hampshire Ave & 24th St NW | 31275 | I Av |

| | duration | start_ts | end_ts | start_station_id | start_station_name | end_station_id | end |
|---------------|----------|---------------------|---------------------|------------------|--------------------|----------------|---------|
| 402193 | 3316 | 2016-11-06 01:30:18 | 2016-11-06 01:25:34 | 31277 | 17th & G St NW | 31289 | He L |

In [18]:

```
# Bei einigen Zeilen ist die Rückgabezeit falsch, dies führt zu Problemen bei der Zählung der Fahrräder
# Korrektur auf Entnahmezeit plus Delta (1 Sekunde)
```

In [19]:

```
df_check.loc[df_check['start_ts'] >= df_check['end_ts'], 'end_ts'] = \
    df_check.loc[df_check['start_ts'] >= df_check['end_ts']]['start_ts'] + pd.Timedelta(1, 'second')
```

In []:

In [20]:

```
# Nochmals prüfen, ob Rückgabezeit stets nach Entnahmezeit
df_check[df_check['start_ts'] >= df_check['end_ts']]
```

Out[20]:

| duration | start_ts | end_ts | start_station_id | start_station_name | end_station_id | end_station_ |
|----------|----------|--------|------------------|--------------------|----------------|--------------|
|----------|----------|--------|------------------|--------------------|----------------|--------------|

In []:

In [21]:

```
# Die Simulation der zeitlich sortierten Entnahmen und Rückgaben (Notebook 25) hat gezeigt,
# dass es Trips gibt, die parallel zu anderen Trips mit demselben Fahrrad stattfinden.
# Diese müssen entfernt werden:
```

In [22]:

```
df_check[(df_check.bike_number=='W00662') & (df_check.start_ts >= pd.Timestamp(2018, 12, 2))][TRIP_COL_LIST_TO_KEEP].head(10)
```

Out[22]:

| start_ts | end_ts | start_station_id | end_station_id | bike_number | Member type |
|----------|--------|------------------|----------------|-------------|-------------|
|----------|--------|------------------|----------------|-------------|-------------|

In [23]:

```
# einige Trips sind nicht möglich (vgl. Simulation)
trips_to_drop = [219942, 220881]
```

In [24]:

```
df_check.drop(trips_to_drop, inplace=True)
```

In []:

In [25]:

```
# Hier Änderung zu seinem Notebook: Mitgliedsstatus wird nicht entfernt, oben in erster
Zeile geändert
```

In [26]:

```
# Beschränkung auf die relevanten Attribute (spart Speicherplatz)
df_trips_clean = df_check[TRIP_COL_LIST_TO_KEEP]
```

In [27]:

```
df_trips_clean.info()
```

```
<class 'pandas.core.frame.DataFrame'>
Int64Index: 10277653 entries, 130487 to 814482
Data columns (total 6 columns):
#   Column                Dtype
---  -
0   start_ts              datetime64[ns]
1   end_ts                datetime64[ns]
2   start_station_id      int64
3   end_station_id        int64
4   bike_number           object
5   Member type           object
dtypes: datetime64[ns](2), int64(2), object(2)
memory usage: 548.9+ MB
```

In []:

In [28]:

```
# Erzeuge zwei separate Merkmale - Tag (date) als timestamp und Stunde (hour) als int
# die spätere Verwendung wie z.B. Gruppierung nach Stunden
df_trips_clean.loc[:, 'start_date'] = df_trips_clean['start_ts'].apply(lambda dt: pd.Timestamp(dt.date()))
df_trips_clean.loc[:, 'start_hour'] = df_trips_clean['start_ts'].apply(lambda dt: dt.time().hour)

df_trips_clean.loc[:, 'end_date'] = df_trips_clean['end_ts'].apply(lambda dt: pd.Timestamp(dt.date()))
df_trips_clean.loc[:, 'end_hour'] = df_trips_clean['end_ts'].apply(lambda dt: dt.time().hour)
```

C:\Users\Simon\anaconda3\lib\site-packages\pandas\core\indexing.py:845: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
self.obj[key] = _infer_fill_value(value)
```

C:\Users\Simon\anaconda3\lib\site-packages\pandas\core\indexing.py:966: SettingWithCopyWarning:

A value is trying to be set on a copy of a slice from a DataFrame.
Try using .loc[row_indexer,col_indexer] = value instead

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/stable/user_guide/indexing.html#returning-a-view-versus-a-copy

```
self.obj[item] = s
```

In [29]:

```
df_trips_clean.head()
```

Out[29]:

| | start_ts | end_ts | start_station_id | end_station_id | bike_number | Member type |
|---------------|------------------------|------------------------|------------------|----------------|-----------------------|-------------|
| 130487 | 2015-10-15 10:58:35 | 2015-10-15 14:57:10 | 31219 | 31634 | (0x0000000074BEBCE4)? | Member |
| 193289 | 2016-10-18 10:54:16 | 2016-10-18 11:19:17 | 31292 | 31292 | (0x0000000074BEBCE4)? | Member |
| 207012 | 2016-10-19 12:20:37 | 2016-10-19 12:32:45 | 31618 | 31618 | (0x0000000074BEBCE4)? | Member |
| 241628 | 2016-10-22 12:07:42 | 2016-10-22 12:26:22 | 31249 | 31249 | (0x0000000074BEBCE4)? | Member |
| 242374 | 2016-10-22 13:01:26 | 2016-10-22 13:30:08 | 31249 | 31249 | (0x0000000074BEBCE4)? | Member |

In []:

In [30]:

```
# gesäuberte Trip-Date speichern  
df_trips_clean.to_pickle(DATA_PATH+CLEAN_TRIPS_FILE)
```

In [31]:

```
df_trips_clean = pd.read_pickle(DATA_PATH+CLEAN_TRIPS_FILE)
```

In []:

In []:

In []: