

Я использую в работе Jupiter Notebook, библиотеки Python Pandas, Matplotlib, Seaborn. Сначала рассмотрим подробно каждую таблицу с данными.

```
drinks.head()
```

	Напиток	Тип
0	Языкосвязывающий лимонный сквош	Безалкогольный
1	Сливочное пиво	Алкольный
2	Шипучий персиковый чай	Безалкогольный
3	Чешуя дракона	Алкольный
4	Эгнот	Безалкогольный

При необходимости сразу переименуем столбцы для удобства работы:

```
drinks = drinks.rename(columns={'Напиток': 'product', 'Тип': 'drink'})
```

Преобразуем строковый тип данных в дату:

```
daily_volume.info()
```

```
<class 'pandas.core.frame.DataFrame'>  
RangeIndex: 3778274 entries, 0 to 3778273  
Data columns (total 5 columns):  
#   Column      Dtype  
---  ---  
0   city        object  
1   id           int64  
2   product     object  
3   date        object  
4   volume_litr float64  
dtypes: float64(1), int64(1), object(3)  
memory usage: 144.1+ MB
```

```
daily_volume['date'] = pd.to_datetime(daily_volume['date'])
```

Определили к какому типу относится каждый напиток, чтобы при необходимости посмотреть на данные в разрезе алкогольный/безалкогольный напиток:

```
daily_volume = daily_volume.merge(drinks, how='left', on=['product'])
```

В таблице с дневными объемами реализации напитков [daily_volume.csv] посмотрим на первый день и последний сбора данных, а также на количество уникальных значений:

```
daily_volume.date.min()
```

```
'1996-01-01'
```

```
daily_volume.date.max()
```

```
'1997-12-31'
```

```
daily_volume.date.nunique()
```

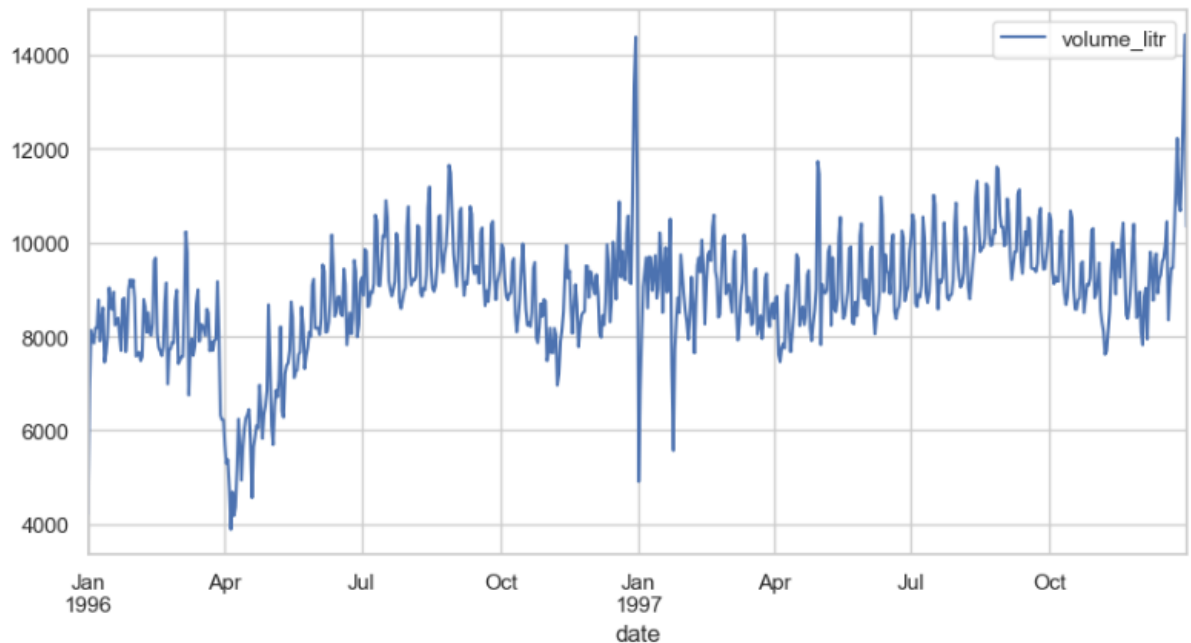
```
731
```

Делаем вывод о том, что данные собирались ежедневно на протяжении двух лет.

Также можем визуализировать дневное потребление для полнейшего представления наших данных (выбрали город Авимор).

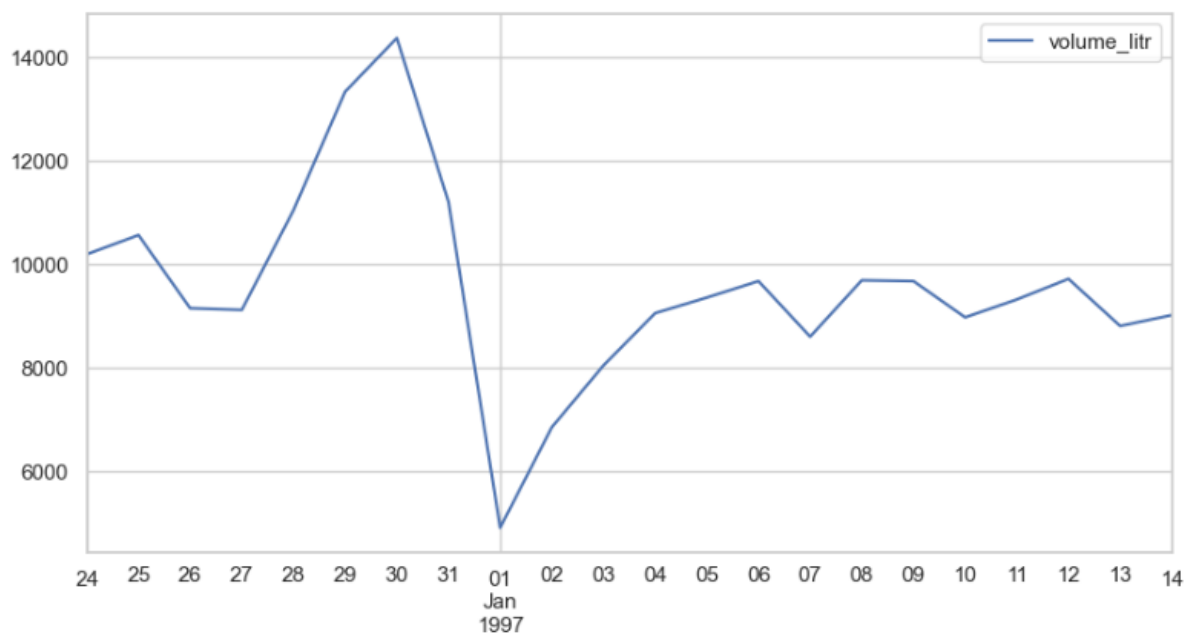
```
sns.set(style='whitegrid', rc={'figure.figsize' : (10,5)})
ax = daily_volume.query('city == "Авимор"' ) \
    .groupby(['date']) \
    .agg({'volume_litr': 'sum'}) \
    .sort_values('date')
ax.plot()
```

<AxesSubplot:xlabel='date'>



```
sns.set(style='whitegrid', rc={'figure.figsize' : (10,5)})
ax = daily_volume.query('city == "Авимор" and date > "1996-12-23" and date < "1997-01-15"' ) \
    .groupby(['date']) \
    .agg({'volume_litr': 'sum'}) \
    .sort_values('date')
ax.plot()
```

<AxesSubplot:xlabel='date'>



Если смотреть по каждому городу, то видим, что в основном максимальные продажи идут за неделю до нового года. Можем подумать, что это влияние праздников, а также накладывается эффект от акции (если акция проводится, далее проверим это).

Рассматривая данные в таблице с эффектами акций [effect.csv] возникают вопросы: как все-таки был посчитан эффект от акции, за какой период идет подсчет. Судя из описания названия столбца это как раз нужный нам для решения поставленной задачи дополнительный объем напитков.

```
effect.date.min()
```

```
Timestamp('1996-01-01 00:00:00')
```

```
effect.date.max()
```

```
Timestamp('1997-12-31 00:00:00')
```

```
effect.date.nunique()
```

```
716
```

Видим, что эффект фиксировался практически ежедневно на протяжении двух лет.

Нашли пропущенные значения в столбце города и по ним же стоят нули в объеме реализации напитков.

```
effect[effect.isnull().values==True]
```

	date	product	city	volume_litr
65	1996-02-10	Безалкогольное	None	0.0
108	1996-12-30	Чешуя дракона	None	0.0
112	1996-08-14	Вода	None	0.0
145	1997-01-03	Сливочное пиво	None	0.0
147	1996-10-04	Вода	None	0.0
...
48879	1997-07-23	Огненный виски	None	0.0
48889	1997-07-24	Чешуя дракона	None	0.0
48897	1997-07-23	Сливочное пиво	None	0.0
48915	1997-07-21	Чешуя дракона	None	0.0
48939	1997-07-24	Смородиновый ром	None	0.0

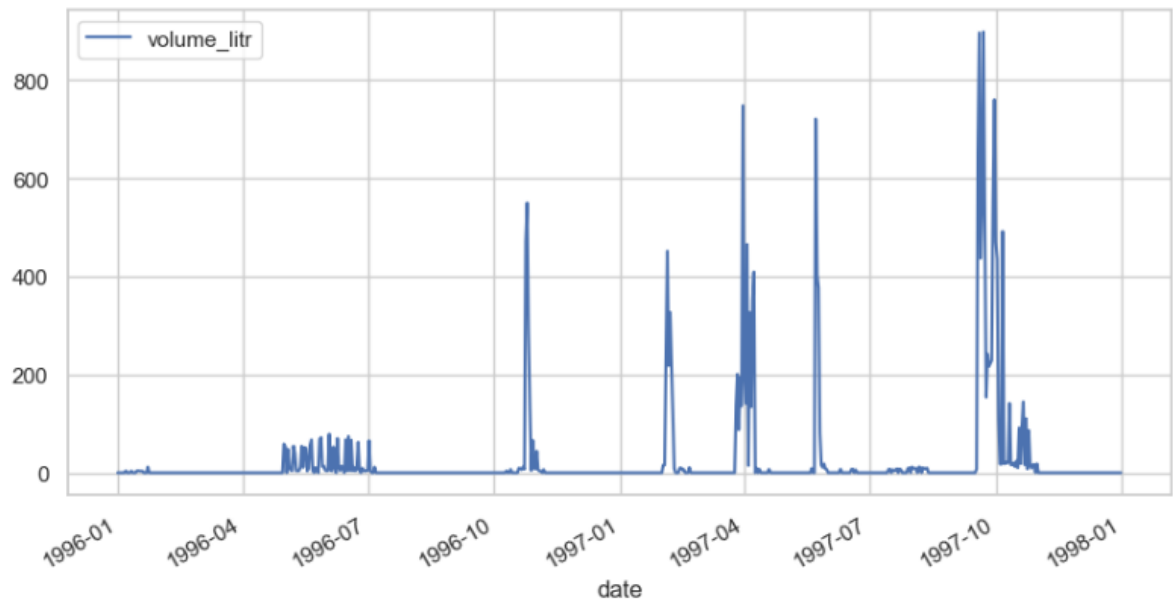
3804 rows x 4 columns

Можем их не удалять, они нам не мешают.

Визуализируем эффект и видим, что на примере города Авимор нет никакого эффекта в новогодние праздники.

```
ax_effect = effect.query('city == "Авимор"' ) \
    .groupby(['date']) \
    .agg({'volume_litr': 'sum'}) \
    .sort_values('date')
ax_effect.plot()
```

<AxesSubplot:xlabel='date'>



Посмотрим одновременно на два графика (график с дневными объемами реализации напитков и график с эффектами акций):

```

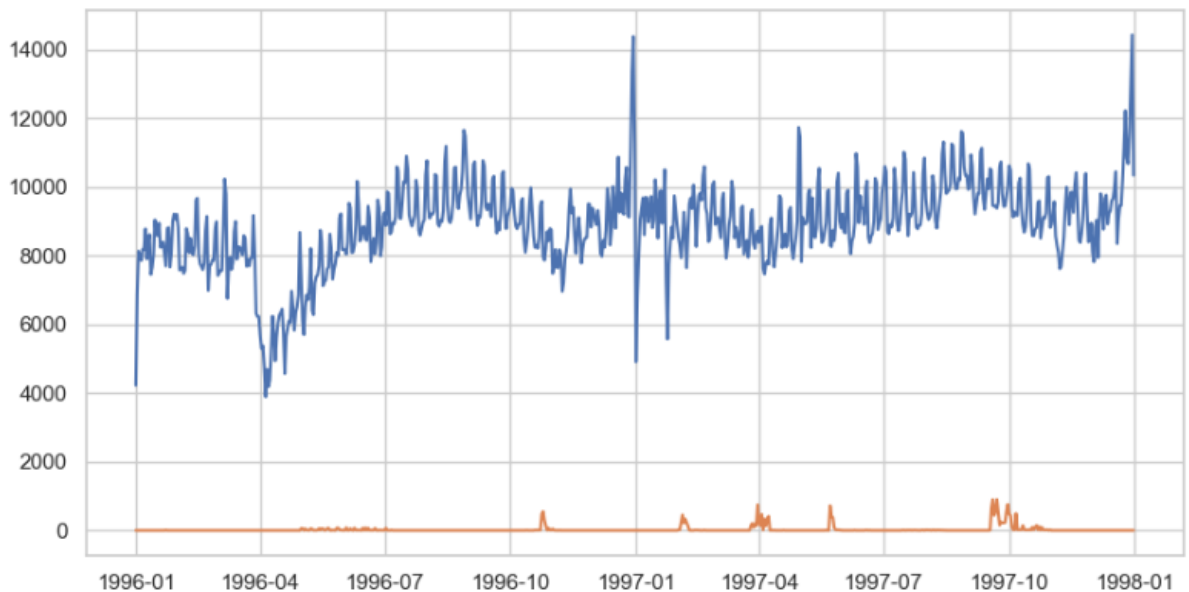
ax_daily_volume = daily_volume.query('city == "Авимор"' ) \
    .groupby(['date']) \
    .agg({'volume_litr': 'sum'}) \
    .sort_values('date')

ax_effect = effect.query('city == "Авимор"' ) \
    .groupby(['date']) \
    .agg({'volume_litr': 'sum'}) \
    .sort_values('date')

pylab.subplot()
pylab.plot(ax_daily_volume)
pylab.plot(ax_effect)

pylab.show()

```



Видим, что почему-то нет корреляции между эффектом и продажами – это странно.

В таблице с реализацией напитков клиентам [sales.csv] также представлены данные за 2 года.

```
sales['date'] = pd.to_datetime(sales['date'])
```

```
sales.date.unique()
```

```
731
```

```
sales.volume_litr.sum()
```

```
1994613.2305577036
```

```
daily_volume.volume_litr.sum()
```

```
82371085.64111032
```

```
sales.segment.unique()
```

```
array(['Орден Феникса', 'Пожиратели смерти'], dtype=object)
```

Смотря на таблицы с дневными объемами реализации напитков [daily_volume.csv] и с реализацией напитков клиентам [sales.csv] могли предположить, что во второй лишь идет разделение по группам всех клиентов, но посмотрев на общий объем в литрах видим существенную разницу, да и всего две группы, поэтому можем сделать вывод о том, что в первой

[daily_volume.csv] реализация всем жителям, даже тем, кто не состоит в группах Ордена Феникса и пожирателей смерти.

В таблице с вероятностью нападения [attack_probability.csv] представлены данные уже не за два года как во всех предыдущих, а всего за 11 месяцев.

```
attack_probability.date.nunique()
```

335

```
attack_probability.date.min()
```

'1996-02-23'

```
attack_probability.date.max()
```

'1997-01-22'

Можем посчитать корреляцию эффекта реализации напитков и вероятности нападения.

```
attack_probability_for_corr = attack_probability.groupby(['city', 'date'], as_index=False) \
    .agg({'attack_probability': 'sum'})
effect_for_corr = effect.groupby(['city', 'date'], as_index=False) \
    .agg({'volume_litr': 'sum'})
corr = attack_probability_for_corr.merge(effect_for_corr, how='left', on=['date', 'city'])
```

```
sns.set(style='whitegrid', rc={'figure.figsize' : (10,5)})
sns.scatterplot(x = 'attack_probability', y = 'volume_litr', data = corr)
plt.title('Взаимосвязь нападения к дополнительному эффекту')
plt.xlabel('Вероятность нападения')
plt.ylabel('Объем эффекта в дополнительных литрах')
```

```
Text(0, 0.5, 'Объем эффекта в дополнительных литрах')
```



Однозначно сложно сказать, что-либо по такой корреляции, так как видим, что при вероятности больше 0,6 дополнительного эффекта почти нет, но также его не наблюдается и при вероятности меньше 0,1.

В таблице с описанием акций [promo.csv] видим, что пропущены 4 названия акции, а также только на 29 акций есть данные по затратам.

После сортировки по началу акции видим, что акции пересекаются! Это не дает нам возможности как-либо распределить наш дополнительный объем. Не хватает состава акции (какие напитки входили в каждую акцию), возможно акции пересекаются с разными напитками.

```
promo = promo.sort_values('Начало')
```

```
promo.head()
```

	Название	Начало	Конец	Кромер	Париж	Бат	Лондон	Салем	Коукворт	Авимор	Даффтаун	Абердин	Нью-Йорк
3	None	1996-01-01	1996-02-29	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0
24	None	1996-01-13	1996-01-26	1.0	1.0	0.0	0.0	1.0	0.0	1.0	1.0	1.0	1.0
25	Минус кнат алкогольное	1996-01-31	1996-02-25	1.0	1.0	1.0	1.0	1.0	0.0	1.0	0.0	0.0	1.0
29	С литрушки на лягушку скидка на безалкогольные...	1996-02-07	1996-03-31	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0	1.0
28	Скидка на Сливочное пиво 2 кната если на барм...	1996-02-07	1996-02-21	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0

Если вернуться к городу Авимор, то в период новогодних праздников акции все же были, но как говорится ранее эффекта от них нет, какая акция не эффективна сказать сложно, так как они пересекаются, но конечно сузить диапазон мы смогли.

```
promo.query('Начало > "1996-12-01" and Начало < "1997-02-01"')
```

	Название	Начало	Конец	Кромер	Париж	Бат	Лондон	Салем	Коукворт	Авимор	Даффтаун	Абердин	Нью-Йорк	Ли Уинг
17	Партнер заведений магазин Ханидьюкс 1997	1996-12-21	1997-05-31	1.0	1.0	1.0	1.0	1.0	1.0	1.0	0.0	1.0	1.0	
63	Акция при поимке Пикси от 15 литров напитков	1996-12-21	1997-01-21	1.0	0.0	1.0	0.0	1.0	1.0	1.0	1.0	0.0	1.0	
41	С литрушки на лягушку	1997-01-01	1997-01-31	1.0	1.0	1.0	0.0	1.0	1.0	1.0	1.0	1.0	1.0	
18	Партнер заведений магазин Ханидьюкс	1997-01-15	1997-05-31	1.0	1.0	0.0	0.0	0.0	1.0	1.0	0.0	0.0	1.0	

Так же интересен тот факт, что если рассмотреть промежутки, например с февраля по май месяц, то видим, что на акции были потрачены деньги, а эффекта практически нет.

Подводя итог, можем сказать, что по имеющимся данным мы не увидели корреляцию между дневными объемами реализации напитков и эффектами акций, также нет результата от акций на которые потрачены средства. Акции между собой пересекаются и отсутствие дополнительной информации не позволяет нам распределить эффект между ними.