

Web scraping & data analysis project

1. Intro
2. Creating the data set
3. Cleaning the data set
4. Exploring patterns
5. Summary

1. Intro

I created a dataset by scraping customer reviews of the Pandora online shop on trustpilot.com. My goal is to identify trends and answer some interesting data questions:

1. Where do the customers who wrote the reviews come from?
2. How satisfied are the customers in general?
3. What's the distribution of the reviews written over the years?
4. Have the star ratings improved over the years?
5. How long do the customers usually wait until they write a review?
6. Is there a correlation between this period of time and the number of stars? Do the unhappy customers write their reviews sooner?

2. Creating the data set

I used several sources that helped me write the code:

https://www.youtube.com/watch?v=HiOtQMcl5wg&ab_channel=AlexTheAnalyst

Angela Yu Python Bootcamp on Udemy

https://www.youtube.com/watch?v=ng2o98k983k&ab_channel=CoreySchafer

ChatGPT

```
In [5]: # importing modules
```

```
In [432]: import requests
from bs4 import BeautifulSoup
import csv
import pandas as pd
import matplotlib.pyplot as plt
import seaborn as sb
```

```
In [ ]: # web scraping through becharming.com
```

```
In [ ]: search_url = 'https://www.trustpilot.com/review/becharming.com?page='
page_nr = 1

while True:

    url = search_url + str(page_nr)
```

```

source = requests.get(url).text

soup = BeautifulSoup(source, 'html.parser')

# print(soup)

articles = soup.find_all('article',
                           class_='paper_paper__1PY90 paper_outline__lwsUX card_car

for article in articles:

    # name of the customer
    name = article.aside.text
    name = name[2:-11]

    print(name)

    # country of the customer
    country = article.aside.text[-2:]

    print(country)

    # star rating

    img_stars = article.img
    # print(img_stars)

    stars = img_stars.get('alt')
    print(stars)

    # review text

    # print(article.prettify())

    review = article.find('p', class_=
    'typography_body-l__KUYFJ typography_appearance-default__AAY17 typography_col

    print(review.string) if review != None else ""

    # print(article.prettify())

    date_experience = article.find('p', class_=
    'typography_body-m__xgxZ_ typography_appearance-default__AAY17')

    date_experience = date_experience.text[20:]

    print(date_experience)

    # date of review
    # print(article.prettify())

    date_review = article.find('time').text
    print(date_review)

    # creating csv file

    data = [name, country, stars, review, date_experience, date_review]

    with open('pandora_reviews.csv', 'a', newline='', encoding='UTF8') as file:
        writer = csv.writer(file)
        writer.writerow(data)

    # pages 1-337

if page_nr < 337:
    page_nr+=1

```

Now the data set is ready. I visited the website on 10 May 2023 and it showed that there are 8358 reviews available, distributed over 337 pages. However, my web scraper was detected and blocked, so that I got 5956 reviews, with the oldest being from 23 February 2016. On the website, the oldest review is from 9 October 2015, so we do not have the data from the first 4,5 months. I will work with almost 6000 reviews from 2016 till 2023, which will surely reveal some interesting trends.

3. Cleaning the data set

There are several problems that have to be fixed in the data:

1. We need to add a header.
2. The star ratings have to be reduced to 1 number (instead of "Rated 5 out of 5 stars" it should only say "5").
3. The code from the review text has to be removed.
4. 2 reviews should have the date instead of "8 hours ago".
5. Both date columns have to be converted to DateTime objects.
6. I want to enrich the data by creating a new column that shows the period of time between the experience and the review in order to use it for further analysis.

```
In [11]: # reading the CSV file with Pandas & creating a copy of the dataframe
```

```
In [433... original_df = pd.read_csv('web_scraper_pandora.csv', header=None)
```

```
In [434... pd.set_option('display.max_colwidth', None)
```

```
In [435... original_df.head(10)
```

0	Donna Stearn	US	Rated 5 out of 5 stars	<p class="typography_body-l_KUYFJ typography_appearance-default__AAY17 typography_color-black__5LYEn" data-service-review-text-typography="true">I have purchased several Pandora pieces from becharming and have always had a great experience. Online ordering is easy and quick and I receive my order quickly.</p>	April 20, 2023	8 hours ago
1	Lashay Irons	US	Rated 5 out of 5 stars	<p class="typography_body-l_KUYFJ typography_appearance-default__AAY17 typography_color-black__5LYEn" data-service-review-text-typography="true">Got on the site, found exactly what I wanted at a reasonable price, and received product in the mail a day or so before expected.</p>	April 07, 2023	Apr 26, 2023
2	Debbie L. Gayhart	US	Rated 5 out of 5 stars	<p class="typography_body-l_KUYFJ typography_appearance-default__AAY17 typography_color-black__5LYEn" data-service-review-text-typography="true">The site is very descriptive. The ring that I ordered looked exactly like the photo and description. I was also very happy with the information regarding shipping. They tell you it will be shipped that day (possibly the next due to time or holiday) and it went out and was delivered quickly. I did not have to pay for any kind of express handling to get it by the date I wanted it.</p>	April 03, 2023	Apr 17, 2023
3	Dee Web	US	Rated 2 out of 5 stars	<p class="typography_body-l_KUYFJ typography_appearance-default__AAY17 typography_color-black__5LYEn" data-service-review-text-typography="true">Item was really small and not what I expected</p>	April 20, 2023	8 hours ago
4	Rosemarie Bieniakowsk	US	Rated 5 out of 5 stars	<p class="typography_body-l_KUYFJ typography_appearance-default__AAY17 typography_color-black__5LYEn" data-service-review-text-typography="true">i received my order in a timely manner . i have total trust in the company and web site. Prices are reasonable and or below cost</p>	March 28, 2023	Apr 27, 2023
5	Patricia Henson	US	Rated 2 out of 5 stars	<p class="typography_body-l_KUYFJ typography_appearance-default__AAY17 typography_color-black__5LYEn" data-service-review-text-typography="true">This is a very cheap piece of jewelry. I realize it's stainless steel but I have other jewelry like that and it is definitely better quality. Very dissatisfied with my purchase.</p>	March 27, 2023	Apr 16, 2023
6	Ann Phillips	US	Rated 5 out of 5 stars	<p class="typography_body-l_KUYFJ typography_appearance-default__AAY17 typography_color-black__5LYEn" data-service-review-text-typography="true">I always enjoy shopping with you! Great Pandora items! Great prices, availability, and prompt shipment! You can't do better than that! "A Satisfied Customer in NC.!"</p>	March 22, 2023	Apr 10, 2023
7	L	US	Rated 5 out of 5 stars	<p class="typography_body-l_KUYFJ typography_appearance-default__AAY17 typography_color-black__5LYEn" data-service-review-text-typography="true">I will definitely buy from you guys again speedy service and great customer service A+ 🌟</p>	April 13, 2023	Apr 27, 2023
8	sonia carne	US	Rated 5 out of 5 stars	<p class="typography_body-l_KUYFJ typography_appearance-default__AAY17 typography_color-black__5LYEn" data-service-review-text-typography="true">My item was shipped fast & when the company made a mistake & sent me the same thing again, they were quick to send me a return shipping label.</p>	March 28, 2023	Apr 13, 2023
9	MG	US	Rated 5 out of 5 stars	<p class="typography_body-l_KUYFJ typography_appearance-default__AAY17 typography_color-black__5LYEn" data-service-review-text-typography="true">I wanted to remove my Pandora bracelet from the original box it came in and place it inside</p>	February 10, 2023	Mar 17, 2023

0 1 2 3 4 5

a safe pouch that I can see through my jewelry box. I was honestly afraid PMOA/BeCharming was not legit because the official Pandora jewelry store I called said they were not affiliated with PMOA/BeCharming.
It wasn't that serious as long as I wasn't ordering from a scammer an could get this cute pouch. Well I placed my order on a Friday afternoon (free shipping option). My order shipped the next day on Saturday (rec'd email notification). My order arrived by USPS on Tuesday. My pouch appears well made and exactly what I needed!</p>

```
In [436... df = original_df.copy()
```

```
In [437... df.head()
```

	0	1	2	3	4	5
				<p class="typography_body-l__KUUFJ typography_appearance-default__AAY17 typography_color-black__5LYEn" data-service-review-text- typography="true">I have purchased several Pandora pieces from becharming and have always had a great experience. Online ordering is easy and quick and I receive my order quickly.</p>	April 20, 2023	8 hours ago
0	Donna Stearn	US	Rated 5 out of 5 stars	<p class="typography_body-l__KUUFJ typography_appearance-default__AAY17 typography_color-black__5LYEn" data-service-review-text- typography="true">Got on the site, found exactly what I wanted at a reasonable price, and received product in the mail a day or so before expected.</p>	April 07, 2023	Apr 26, 2023
1	Lashay Irons	US	Rated 5 out of 5 stars	<p class="typography_body-l__KUUFJ typography_appearance-default__AAY17 typography_color-black__5LYEn" data-service-review-text- typography="true">The site is very descriptive. The ring that I ordered looked exactly like the photo and description. I was also very happy with the information regarding shipping. They tell you it will be shipped that day (possibly the next due to time or holiday) and it went out and was delivered quickly. I did not have to pay for any kind of express handling to get it by the date I wanted it.</p>	April 03, 2023	Apr 17, 2023
2	Debbie L. Gayhart	US	Rated 5 out of 5 stars	<p class="typography_body-l__KUUFJ typography_appearance-default__AAY17 typography_color-black__5LYEn" data-service-review-text- typography="true">Item was really small and not what I expected</p>	April 20, 2023	8 hours ago
3	Dee Web	US	Rated 2 out of 5 stars	<p class="typography_body-l__KUUFJ typography_appearance-default__AAY17 typography_color-black__5LYEn" data-service-review-text- typography="true">i received my order in a timely mannor . i have total trust in the company and web site. Prices arereasonable and or below cost</p>	March 28, 2023	Apr 27, 2023
4	Rosemarie Bieniakowsk	US	Rated 5 out of 5 stars			

```
In [438... # adding a header
```

```
In [439... df.to_csv('pandora_reviews.csv',
            header=['name', 'country_origin', 'number_stars', 'review_text', 'date_expe
            index=False)
```

```
In [440... df = pd.read_csv('pandora_reviews.csv')
```

```
In [441... df.head()
```

Out [441] :

	name	country_origin	number_stars	review_text	date_experience	date_review
0	Donna Stearn	US	Rated 5 out of 5 stars	<p><p class="typography_body-I_KUYFJ typography_appearance-default__AAY17 typography_color-black__5LYEn" data-service-review-text-typography="true">I have purchased several Pandora pieces from becharming and have always had a great experience. Online ordering is easy and quick and I receive my order quickly.</p></p>	April 20, 2023	8 hours ago
1	Lashay Irons	US	Rated 5 out of 5 stars	<p><p class="typography_body-I_KUYFJ typography_appearance-default__AAY17 typography_color-black__5LYEn" data-service-review-text-typography="true">Got on the site, found exactly what I wanted at a reasonable price, and received product in the mail a day or so before expected.</p></p>	April 07, 2023	Apr 26, 2023
2	Debbie L. Gayhart	US	Rated 5 out of 5 stars	<p><p class="typography_body-I_KUYFJ typography_appearance-default__AAY17 typography_color-black__5LYEn" data-service-review-text-typography="true">The site is very descriptive. The ring that I ordered looked exactly like the photo and description. I was also very happy with the information regarding shipping. They tell you it will be shipped that day (possibly the next due to time or holiday) and it went out and was delivered quickly. I did not have to pay for any kind of express handling to get it by the date I wanted it.</p></p>	April 03, 2023	Apr 17, 2023
3	Dee Web	US	Rated 2 out of 5 stars	<p><p class="typography_body-I_KUYFJ typography_appearance-default__AAY17 typography_color-black__5LYEn" data-service-review-text-typography="true">Item was really small and not what I expected</p></p>	April 20, 2023	8 hours ago

	name	country_origin	number_stars	review_text	date_experience	date_review
4	Rosemarie Bieniakowsk	US	Rated 5 out of 5 stars	<div> <div> <div></div> <div></div> <div></div> <div></div> <div></div> </div> <div> <div></div> <div></div> <div></div> <div></div> <div></div> </div> </div> <p> received my order in a timely manner . i have total trust in the company and web site. Prices are reasonable and or below cost </p>	March 28, 2023	Apr 27, 2023

In [442...


```
# general information
```

In [443...


```
df.describe()
```

Out[443]:

	name	country_origin	number_stars	review_text	date_experience	date_review
count	5937	5940	5817	5214	5940	5940
unique	5590	49	5	5205	1922	2013

Great service

top	customer	US	Rated 5 out of 5 stars	December 11, 2017	Dec 11, 2017
-----	----------	----	------------------------	-------------------	--------------

freq	15	5744	5033	3	50	46
------	----	------	------	---	----	----

In [444...


```
df.info()
```

```

<class 'pandas.core.frame.DataFrame'>
RangeIndex: 5940 entries, 0 to 5939
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  -
0   name                   5937 non-null   object
1   country_origin         5940 non-null   object
2   number_stars           5817 non-null   object
3   review_text            5214 non-null   object
4   date_experience         5940 non-null   object
5   date_review            5940 non-null   object
dtypes: object(6)
memory usage: 278.6+ KB

```

In [445...


```
# changing the format of the stars column
```

In [446...


```
df['number_stars'].describe()
```

Out[446]:

count	5817
unique	5
top	Rated 5 out of 5 stars
freq	5033
Name:	number_stars, dtype: object

In [447...


```
df['number_stars'].isnull().sum()
```

Out[447]: 123

```
In [448... x = "Rated 5 out of 5 stars"
x = x[6]
x
```

Out[448]: '5'

```
In [449... if not df['number_stars'].isnull().all():
    df['number_stars'] = df['number_stars'].str.extract('(\d)')
```

```
In [450... df['number_stars'].sample(10)
```

```
Out[450]: 4038    5
1528    5
147     5
1952    5
3998    5
5753    5
5163    4
1982    5
1513    5
3823    5
Name: number_stars, dtype: object
```

```
In [451... # converting to integers
```

```
In [452... df['number_stars'] = pd.to_numeric(df['number_stars'], errors='coerce').astype('Int64
```

```
In [453... df['number_stars'].info()
```

```
<class 'pandas.core.series.Series'>
RangeIndex: 5940 entries, 0 to 5939
Series name: number_stars
Non-Null Count  Dtype
-----
5817 non-null   Int64
dtypes: Int64(1)
memory usage: 52.3 KB
```

```
In [454... # fixing the review text
```

```
In [455... df['review_text'].head()
```



```

Out[455]: 0
<p class="typography_body-1__KUYFJ typography_appearance-default__AAY17 typography_color-black__5LYEn" data-service-review-text-typography="true">I have purchased several Pandora pieces from becharming and have always had a great experience. Online ordering is easy and quick and I receive my order quickly.</p>
1
<p class="typography_body-1__KUYFJ typography_appearance-default__AAY17 typography_color-black__5LYEn" data-service-review-text-typography="true">Got on the site, found exactly what I wanted at a reasonable price, and received product in the mail a day or so before expected.</p>
2    <p class="typography_body-1__KUYFJ typography_appearance-default__AAY17 typography_color-black__5LYEn" data-service-review-text-typography="true">The site is very descriptive. The ring that I ordered looked exactly like the photo and description. I was also very happy with the information regarding shipping. They tell you it will be shipped that day (possibly the next due to time or holiday) and it went out and was delivered quickly. I did not have to pay for any kind of express handling to get it by the date I wanted it.</p>
3
<p class="typography_body-1__KUYFJ typography_appearance-default__AAY17 typography_color-black__5LYEn" data-service-review-text-typography="true">Item was really small and not what I expected</p>
4
<p class="typography_body-1__KUYFJ typography_appearance-default__AAY17 typography_color-black__5LYEn" data-service-review-text-typography="true">i received my order in a timely mannor . i have total trust in the company and web site. Prices are reasonable and or below cost</p>
Name: review_text, dtype: object

```

```

In [456... df['review_text'] = df['review_text'].str.replace('<p class="typography_body-1__KUYFJ

```

```

In [457... df['review_text'].head()

```

```

Out[457]: 0
I have purchased several Pandora pieces from becharming and have always had a great experience. Online ordering is easy and quick and I receive my order quickly.</p>
1
Got on the site, found exactly what I wanted at a reasonable price, and received product in the mail a day or so before expected.</p>
2    The site is very descriptive. The ring that I ordered looked exactly like the photo and description. I was also very happy with the information regarding shipping. They tell you it will be shipped that day (possibly the next due to time or holiday) and it went out and was delivered quickly. I did not have to pay for any kind of express handling to get it by the date I wanted it.</p>
3
Item was really small and not what I expected</p>
4
i received my order in a timely mannor . i have total trust in the company and web site. Prices are reasonable and or below cost</p>
Name: review_text, dtype: object

```

```

In [458... df['review_text'] = df['review_text'].str.replace('</p>','')

```

```

In [459... df['review_text'] = df['review_text'].str.replace('<br/>','')

```

```

In [460... df['review_text'].sample(10)

```

```
Out[460]: 536
NaN
5104
From the online shopping to receiving my purchase, the entire experience was easy &
mp; professional. When my safety chain arrive speedier than expected, I was extremel
y pleased with the look and fit of this item.
126
Item was delivered quickly and in perfect condition.
5883
I can't say enough nice things about this company. It is so very easy to order onlin
e, they ship right away and usually arrive before the delivery date you are given.
Every piece comes in its own box to make it easy for gift giving. Not some cheap fo
ld up paper box either. I highly recommend this company. I've ordered four times n
ow and am a Very Happy Customer!!!
1796
NaN
2603
The item I ordered was received quickly and the process was easy.
1723 I tried to place an order and my payment was declined so I assumed the order
didn't go through. The next day I had confirmation of my order so I emailed straight
back to say I wanted to cancel the order as I did not wish to proceed and would like
a full refund. I received an email back asking me to reconsider and I advised that I
still wanted a refund. I received an email back to confirm that my refund had been p
rocessed. I have yet to receive my refund and it was weeks ago and my emails are not
being responded to. So basically I have paid for something that I have not received.
4238
I do not have any problemsvery happy with purchase
171
fast shipping
1043
First time shopper & amazed at the fast service. Item arrived as described and b
eautifully packaged. Highly recommend this seller.
Name: review_text, dtype: object
```

```
In [461... # changing the date_review from hours to the date
```

```
In [462... df['date_review'] = df['date_review'].str.replace('8 hours ago','May 10, 2023')
```

```
In [463... df['date_review'].head()
```

```
Out[463]: 0    May 10, 2023
1    Apr 26, 2023
2    Apr 17, 2023
3    May 10, 2023
4    Apr 27, 2023
Name: date_review, dtype: object
```

```
In [464... # converting the date_experience to a date time object
```

```
In [465... df['date_experience'] = pd.to_datetime(df['date_experience'])
```

```
In [466... df['date_experience'].head()
```

```
Out[466]: 0    2023-04-20
1    2023-04-07
2    2023-04-03
3    2023-04-20
4    2023-03-28
Name: date_experience, dtype: datetime64[ns]
```

```
In [467... # converting the date_review to a date time object
```

```
In [468... df['date_review'].sample(10)
```

```
Out[468]: 4943    Feb 27, 2017
          4729    Apr 13, 2017
          5910    Feb 29, 2016
          3197    Jan 21, 2019
          1953     Jun 7, 2020
          3014    May 21, 2019
          5168    Jan 30, 2017
          2151    May 17, 2020
          4088    Dec 16, 2017
          1125     Feb 3, 2021
          Name: date_review, dtype: object
```

```
In [469... df['date_review'] = pd.to_datetime(df['date_review'].str.replace('Updated on ', ''),
```

```
In [470... df['date_review'].sample(10)
```

```
Out[470]: 231      2022-11-15
          1079    2021-02-19
          2930    2019-07-15
          5008    2017-02-20
          5879    2016-03-17
          2353    2020-04-01
          3687    2018-06-03
           1     2023-04-26
          1059    2021-02-26
          5924    2016-02-25
          Name: date_review, dtype: datetime64[ns]
```

```
In [471... # creating a new column showing the time difference between the review and the experi
```

```
In [472... df['time_betw_exp_review'] = df['date_review'] - df['date_experience']
```

```
In [473... df.sample(10)
```

	name	country_origin	number_stars	review_text	date_experience	date_review	time_betw
				Becharming is the first place I go to when looking for new charms. I know I can trust them. I've been shopping with them for years.	2021-08-24	2021-08-24	
772	JuliBell1	US	5				
5038	Victor Arzol	US	5	Definitely shop again. Customer support, 2 thumbs up.	2017-02-17	2017-02-17	
5321	JACOB BRESHEAR	US	5	Easy, fast, reliable processing and turnaround of online orders. I come back every year.	2017-01-01	2017-01-01	
4346	geri	US	5	I have bought 11 charms from you and each one more beautiful then the last. I was able to track my package from when it left to when it reached my post office. I will be ordering more charms from you:)	2017-07-26	2017-07-26	
3050	Patty	US	5	I love ordering from Be Charming! I have always received my order in a timely manner and have been truly pleased with the products!	2019-04-29	2019-04-29	
1224	Chris	US	5	NaN	2021-01-07	2021-01-07	
1074	Katina Smit	US	5	Love my purchase and product	2021-02-20	2021-02-20	

	name	country_origin	number_stars	review_text	date_experience	date_review	time_betw
				arrive on time. Will definitely be ordering from this company again.			
2746	D	US	5	Goods were Ordered, Shipped and Received as Expected	2019-12-13	2019-12-13	
825	Cinitia Robert	US	5	Fast and super efficient service. Had minor issue that was resolved quickly. Happy customer!	2021-06-29	2021-06-29	
5426	Deborah P	US	5	I have been an online customer of Be Charming for several years. The site consistently maintains a good selection and inventory of Pandora products. My orders are always timely received, and I have never had an issue that required a call into customer service. To me, this is the best online shopping experience one could ask for!	2016-11-14	2016-11-14	

4. Exploring patterns

1. Where do the customers who wrote the reviews come from?

```
In [474... df.country_origin.describe()
```

```
Out[474]: count      5940  
          unique       49  
          top         US  
          freq      5744  
          Name: country_origin, dtype: object
```

```
In [475]: df.country_origin.value_counts()
```

```
Out[475]: US      5744  
          GB        65  
          PR        23  
          KR        13  
          MX        11  
          CL         5  
          PH         5  
          PA         5  
          NI         4  
          TT         4  
          CA         4  
          JM         4  
          DE         4  
          DO         3  
          BS         3  
          VI         3  
          IT         3  
          IE         2  
          AE         2  
          AL         2  
          JP         2  
          PL         2  
          LC         1  
          AU         1  
          BM         1  
          AM         1  
          SG         1  
          RS         1  
          KN         1  
          PE         1  
          AG         1  
          SV         1  
          BE         1  
          CO         1  
          EC         1  
          CN         1  
          DM         1  
          BR         1  
          UY         1  
          KH         1  
          ZA         1  
          MU         1  
          IR         1  
          AT         1  
          GT         1  
          NL         1  
          NZ         1  
          BW         1  
          ES         1  
          Name: country_origin, dtype: int64
```

The vast majority of customers come from the US (5744 out of 5940), however, there are customer reviews from all over the world.

1. How satisfied are the customers in general?

```
In [476]: df.number_stars.mean()
```

Out [476]: 4.695891352931064

In [477... `df.head()`

Out [477] :

	name	country_origin	number_stars	review_text	date_experience	date_review	time_betw_
0	Donna Stearn	US	5	I have purchased several Pandora pieces from becharming and have always had a great experience. Online ordering is easy and quick and I receive my order quickly.	2023-04-20	2023-05-10	
1	Lashay Irons	US	5	Got on the site, found exactly what I wanted at a reasonable price, and received product in the mail a day or so before expected.	2023-04-07	2023-04-26	
2	Debbie L. Gayhart	US	5	The site is very descriptive. The ring that I ordered looked exactly like the photo and description. I was also very happy with the information regarding shipping. They tell you it will be shipped that day (possibly the next due to time or holiday) and it went out and was delivered quickly. I did not have to pay for any kind of express handling to get it by the date I wanted it.	2023-04-03	2023-04-17	
3	Dee Web	US	2	Item was really small and not what I expected	2023-04-20	2023-05-10	
4	Rosemarie Bieniakowsk	US	5	i received my order in a timely mannor . i	2023-03-28	2023-04-27	

name	country_origin	number_stars	review_text	date_experience	date_review	time_betw_
			have total trust in the company and web site. Prices are reasonable and or below cost			

The overall average number of stars is almost 4.7 out of 5, which reflects a very positive customer experience.

1. Is there a correlation between the date of experience and the number of stars? Has the customer experience of the pandora shoppers improved over time?

```
In [478... df.number_stars.info()

<class 'pandas.core.series.Series'>
RangeIndex: 5940 entries, 0 to 5939
Series name: number_stars
Non-Null Count  Dtype
-----
5817 non-null   Int64
dtypes: Int64(1)
memory usage: 52.3 KB
```

Since there are several missing values, we have to get rid of them before creating a plot.

```
In [479... df_copy = df[df['number_stars'].notnull()].copy()
```

```
In [480... df_copy.number_stars.info()

<class 'pandas.core.series.Series'>
Int64Index: 5817 entries, 0 to 5939
Series name: number_stars
Non-Null Count  Dtype
-----
5817 non-null   Int64
dtypes: Int64(1)
memory usage: 96.6 KB
```

```
In [481... # adding a year column
```

```
In [482... df_copy['year'] = df_copy.date_experience.dt.year
```

```
In [483... df_copy.head()
```

	name	country_origin	number_stars	review_text	date_experience	date_review	time_betw_
0	Donna Stearn	US	5	I have purchased several Pandora pieces from becharming and have always had a great experience. Online ordering is easy and quick and I receive my order quickly.	2023-04-20	2023-05-10	
1	Lashay Irons	US	5	Got on the site, found exactly what I wanted at a reasonable price, and received product in the mail a day or so before expected.	2023-04-07	2023-04-26	
2	Debbie L. Gayhart	US	5	The site is very descriptive. The ring that I ordered looked exactly like the photo and description. I was also very happy with the information regarding shipping. They tell you it will be shipped that day (possibly the next due to time or holiday) and it went out and was delivered quickly. I did not have to pay for any kind of express handling to get it by the date I wanted it.	2023-04-03	2023-04-17	
3	Dee Web	US	2	Item was really small and not what I expected	2023-04-20	2023-05-10	
4	Rosemarie Bieniakowsk	US	5	i received my order in a timely mannor . i	2023-03-28	2023-04-27	

name	country_origin	number_stars	review_text	date_experience	date_review	time_betw_
			have total trust in the company and web site. Prices are reasonable and or below cost			

```
In [484]: df_copy.year.value_counts()
```

```
Out[484]: 2017    1315
          2020    1265
          2018     700
          2021     699
          2019     685
          2016     609
          2022     435
          2023     109
          Name: year, dtype: int64
```

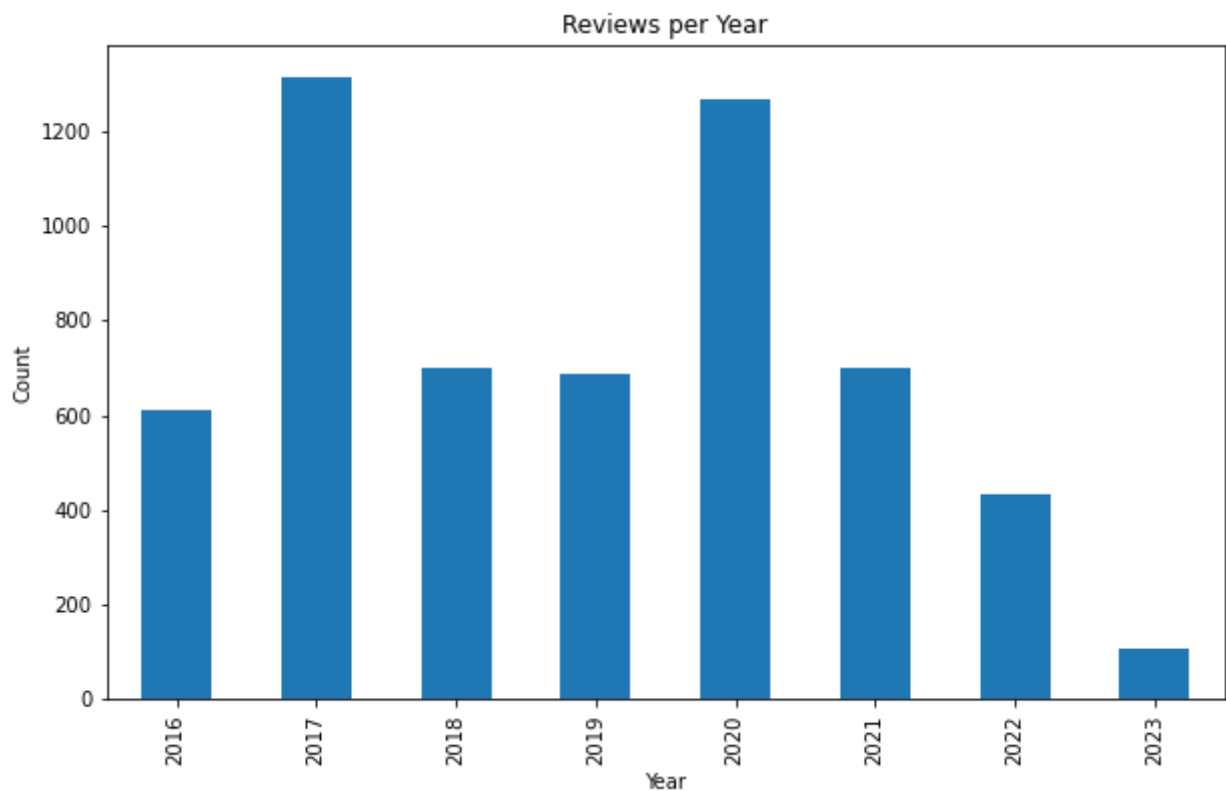
```
In [485]: # visualizing the review counts over the years
```

```
In [486]: year_counts = df_copy['year'].value_counts().sort_index()

plt.figure(figsize=(10, 6))
year_counts.plot(kind='bar')

plt.xlabel('Year')
plt.ylabel('Count')
plt.title('Reviews per Year')

plt.show()
```



We see that most reviews were written in 2017 and in 2020. I do not have enough information on the development of the online shop, their marketing etc. to explain the distribution of the data. I can imagine that the peak in 2020 may be explained by corona, which became a pandemic in 2020 and caused the people to do more online shopping and spend more time in front of the computer because of the lockdowns.

1. Have the star ratings improved over the years?

```
In [487... # plotting the distribution of star ratings over the years
```

```
In [488... def hist(year):
    df_filtered = df_copy[df_copy['year'] == year]
    ax.hist(data=df_filtered, x='number_stars')
    ax.set_xlabel('Star Ratings')
    ax.set_ylabel('Count')
    ax.set_title(f'Distribution of Star Ratings in {year}')

num_rows = 2
num_cols = 4

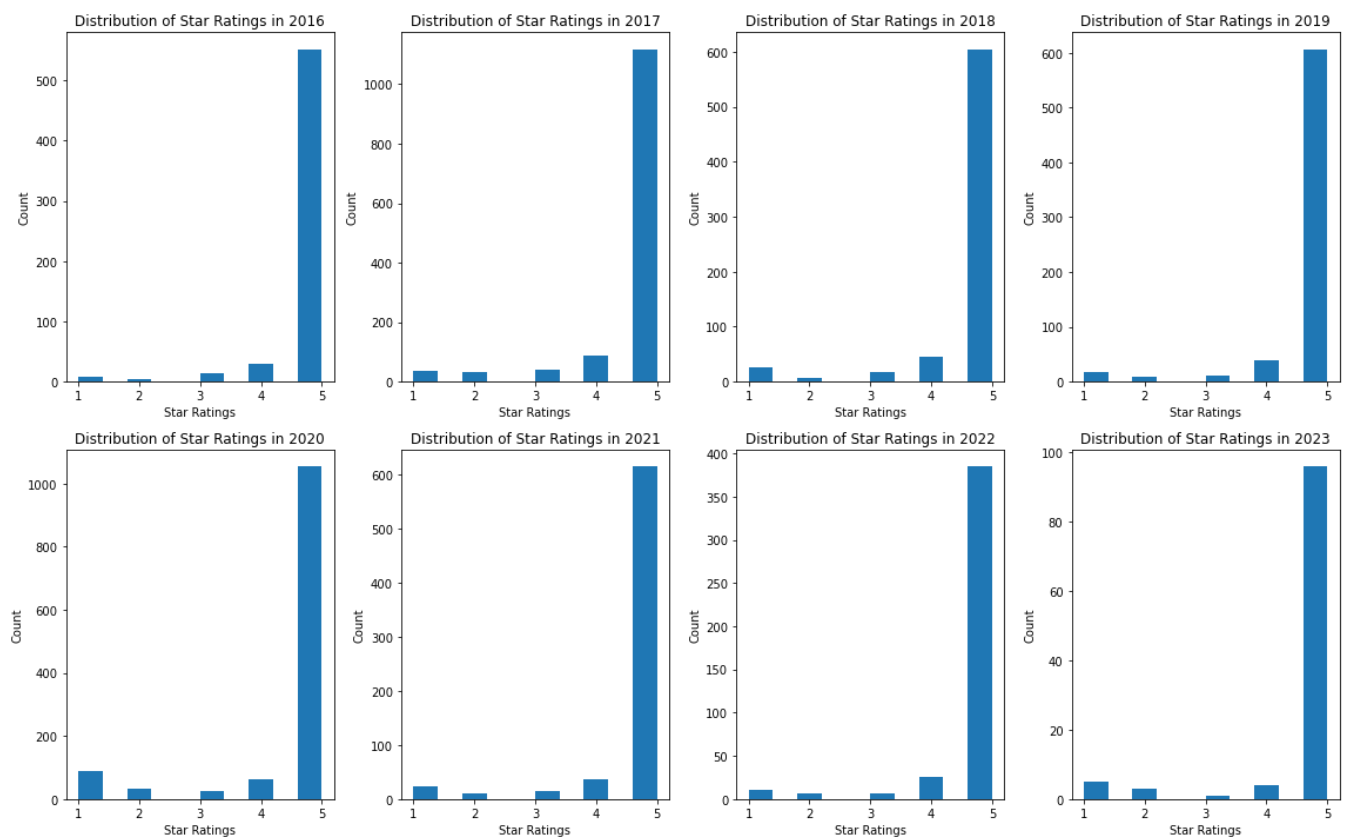
fig, axes = plt.subplots(num_rows, num_cols, figsize=(16, 10))

axes = axes.flatten()

for i, year in enumerate([2016, 2017, 2018, 2019, 2020, 2021, 2022, 2023]):
    ax = axes[i]
    hist(year)

fig.tight_layout()

plt.show()
```



We can observe that there are only fine differences between the star rating distribution over the years. Most customers are very satisfied and have given the online shop 5 stars over the last years.

1. How long do the customers usually wait until they write a review?

```
In [489... df_copy.time_betw_exp_review.mean()
```

```
Out[489]: Timedelta('0 days 16:30:06.623444892')
```

```
In [490... df_copy.time_betw_exp_review.value_counts()
```

```

Out[490]:
0 days      5483
19 days      27
14 days      23
20 days      21
15 days      19
12 days      11
9 days       10
18 days      10
1 days        9
17 days        9
21 days        9
13 days        9
16 days        6
11 days        6
22 days        5
8 days         5
6 days         4
4 days         4
23 days        4
3 days         4
2 days         3
25 days        2
5 days         2
24 days        2
7 days         2
26 days        2
30 days        2
10 days        2
37 days        1
35 days        1
43 days        1
46 days        1
41 days        1
34 days        1
31 days        1
32 days        1
118 days       1
314 days       1
29 days        1
38 days        1
Name: time_betw_exp_review, dtype: int64

```

```
In [492]: df_copy.time_betw_exp_review.max()
```

```
Out[492]: Timedelta('314 days 00:00:00')
```

Most customers write the review the same day they get the product. If the data is accurate and not based on a typo or misunderstanding, the longest time a customer waited to write a review was almost a year (314 days).

1. Is there a correlation between this period of time and the number of stars? Do the unhappy customers write their reviews sooner?

```
In [494]: filtered_df = df_copy[df_copy['number_stars'].isin([1, 2])]
          filtered_time_betw_exp_review = filtered_df['time_betw_exp_review']
```

```
In [495]: filtered_time_betw_exp_review.describe()
```

```
Out [495]: count          297
          mean      0 days 10:44:50.909090909
          std       2 days 18:21:59.987634773
          min              0 days 00:00:00
          25%              0 days 00:00:00
          50%              0 days 00:00:00
          75%              0 days 00:00:00
          max       23 days 00:00:00
          Name: time_betw_exp_review, dtype: object
```

```
In [496... filtered_df2 = df_copy[df_copy['number_stars'].isin([4, 5])]
          filtered_time_betw_exp_review2 = filtered_df2['time_betw_exp_review']
```

```
In [497... filtered_time_betw_exp_review2.describe()
```

```
Out [497]: count          5285
          mean      0 days 17:10:28.722800378
          std       5 days 18:07:48.011302076
          min              0 days 00:00:00
          25%              0 days 00:00:00
          50%              0 days 00:00:00
          75%              0 days 00:00:00
          max      314 days 00:00:00
          Name: time_betw_exp_review, dtype: object
```

It seems that, on average, both the happy customers, who are the majority, as well as the unhappy customers tend to write their reviews the same day.

It would be interesting to look at the distribution of the 5-star reviews over the year and to see if there is a particular season when the customers order more (and write a review, which mostly happens the same day).

```
In [498... # adding a month column to the data frame
```

```
In [499... df_copy['month'] = df_copy.date_experience.dt.month
```

```
In [500... df_copy.head()
```

Out [500] :

	name	country_origin	number_stars	review_text	date_experience	date_review	time_betw_
0	Donna Stearn	US	5	I have purchased several Pandora pieces from becharming and have always had a great experience. Online ordering is easy and quick and I receive my order quickly.	2023-04-20	2023-05-10	
1	Lashay Irons	US	5	Got on the site, found exactly what I wanted at a reasonable price, and received product in the mail a day or so before expected.	2023-04-07	2023-04-26	
2	Debbie L. Gayhart	US	5	The site is very descriptive. The ring that I ordered looked exactly like the photo and description. I was also very happy with the information regarding shipping. They tell you it will be shipped that day (possibly the next due to time or holiday) and it went out and was delivered quickly. I did not have to pay for any kind of express handling to get it by the date I wanted it.	2023-04-03	2023-04-17	
3	Dee Web	US	2	Item was really small and not what I expected	2023-04-20	2023-05-10	
4	Rosemarie Bieniakowsk	US	5	i received my order in a timely mannor . i	2023-03-28	2023-04-27	

name	country_origin	number_stars	review_text	date_experience	date_review	time_betw_
			have total			
			trust in the			
			company and			
			web site.			
			Prices			
			are reasonable			
			and or below			
			cost			

In [501]...

```
# Create a figure and axis
fig, ax = plt.subplots(figsize=(12, 8))

# Iterate through the years
for year in range(2016, 2024):
    # Filter the DataFrame for 5-star reviews in the current year
    filtered_df = df_copy[(df_copy['number_stars'] == 5) & (df_copy['year'] == year)]

    # Count the number of reviews per month
    monthly_counts = filtered_df['month'].value_counts().sort_index()

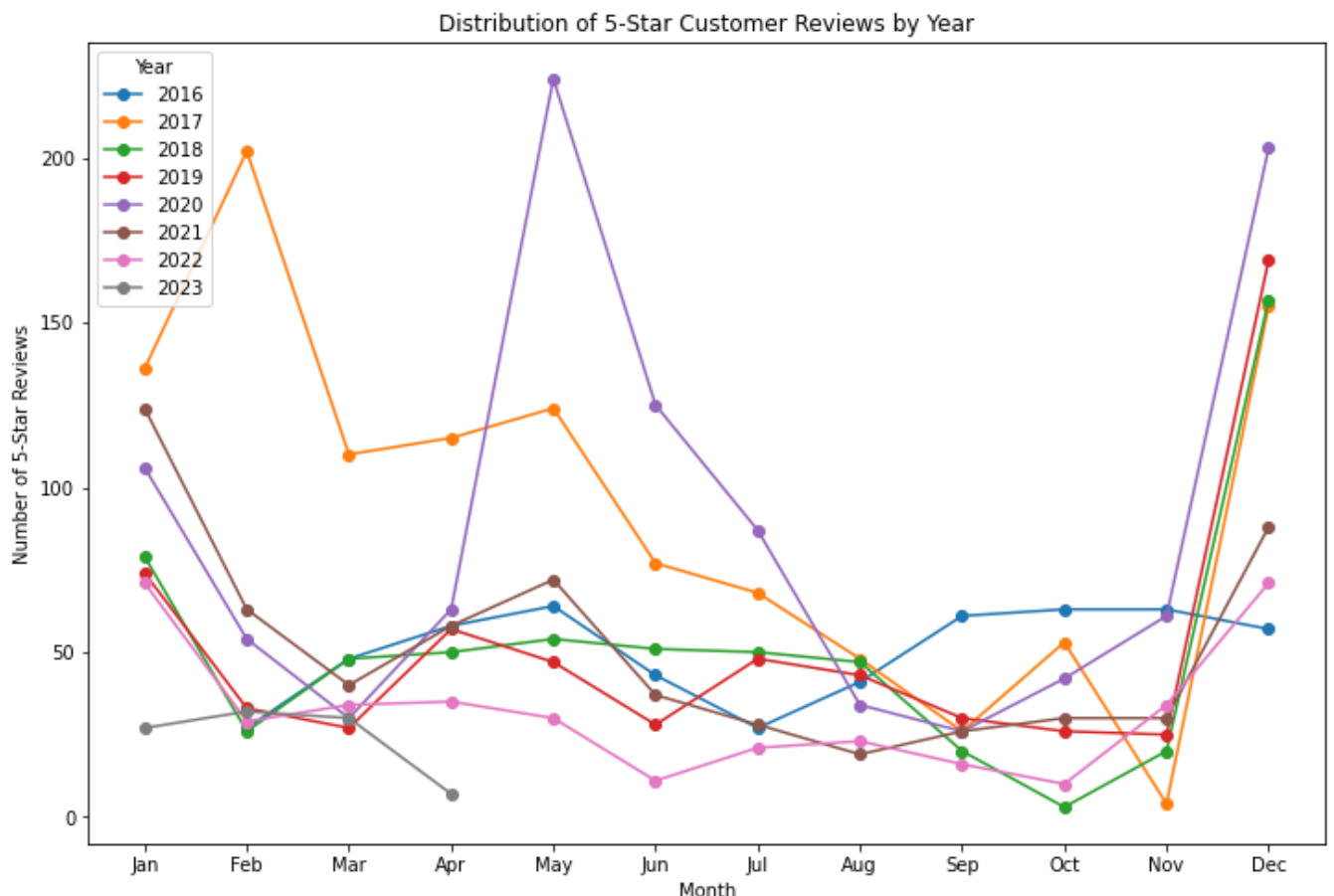
    # Plot the line chart for the current year
    ax.plot(monthly_counts.index, monthly_counts.values, marker='o', linestyle='-', 1

# Set the x-ticks and labels
plt.xticks(range(1, 13), ['Jan', 'Feb', 'Mar', 'Apr', 'May', 'Jun', 'Jul', 'Aug', 'Se

# Set the labels and title
plt.xlabel('Month')
plt.ylabel('Number of 5-Star Reviews')
plt.title('Distribution of 5-Star Customer Reviews by Year')

# Add a legend
ax.legend(title='Year', loc='upper left')

# Show the plot
plt.show()
```



We can see that the month with the highest number of 5-star reviews was May 2020. December 2020 is also a peak, so I assume that both numbers are connected to the holidays in December & Mother's Day in May. We can see a rise in December for all the years apart from 2016. 2017 there was a major peak in February, but then the number falls and goes up again fast in December, which is the season when people buy gifts. The lowest numbers of 5-star reviews can be observed in 2022 and 2023, which did not start in a very promising way. Now let's zoom in on the peaks of May 2020 and February 2017.

```
In [408... df_may2020 = df_copy[(df_copy['month']==5) & (df_copy['year']==2020)]
```

```
In [424... df_may2020.sample(20)
```

	name	country_origin	number_stars	review_text	date_experience	date_review	time_t
2102	Marlene Collazo	US	5	I LOVE BUYING FROM THIS COMPANY! THEY ARE THE BEST SELLING PANDORA STORE I EVER DEALT WITH!!!!HIGHLY RECOMMEND THEM.	2020-05-21	2020-05-21	
2151	Kelsey Harri	US	5	My mom love her bracelet, it's a little small. My opinion is get the 8inch and up for adults. Definitely will be coming back !!! Xoxo	2020-05-17	2020-05-17	
2222	Yanira Elena Davidson	US	5	The care taken into the packaging of my order was felt when I opened it! Amazing working with you guys!!	2020-05-12	2020-05-12	
2271	Debra Pritchett	US	5	Exactly what I wanted - on time - delighted with purchase! Cutest little graduation owl charm for my granddaughter's new Pandora bracelet.	2020-05-04	2020-05-04	
2211	dawn durham	US	5	Everything was great! I had a bracelet and 3 charms delivered to my Godmother in another state. Unfortunately, her bracelet is a tad too small. But the delivery was awesome.	2020-05-13	2020-05-13	
2283	Daisy jane Ro	US	5	NaN	2020-05-01	2020-05-01	
2184	brandon perkerso	US	2	I'm very annoyed that I asked for this to be sent incognito so my wife wouldn't see the label. But thats too much to ask for i suppose especially on mothers day. Thanks for that.	2020-05-15	2020-05-15	

	name	country_origin	number_stars	review_text	date_experience	date_review	time_l
2003	Benita	US	5	Excellence service. Got my shipment quick.	2020-05-29	2020-05-29	
2085	Roberta Warne	US	5	Good selection of merchandise. Fast and timely delivery. Would recommend them to friends and I will continue to buy jewelry from here	2020-05-21	2020-05-21	
2070	Datrina lander	US	5	My mother loved her mother day gift also my grandmother so all in all we all were happy and i was very thrilled about how thry turned out we loved it also will order anytime from this site	2020-05-22	2020-05-22	
2130	Amanda	US	5	NaN	2020-05-19	2020-05-19	
2263	Eleano	US	5	Becharming is a great way to shop for Pandora! Product arrived right on time and as ordered.	2020-05-06	2020-05-06	
2113	Fred Rodrigue	US	5	Easy , quick, and fast	2020-05-20	2020-05-20	
2052	Katherine Clar	US	5	No complaints about using the site, ordering, shipping. All went very well.	2020-05-23	2020-05-23	
2147	Deborah	US	5	Absolutely love my new bracelet. The whole process was streamlined, I will be ordering more!	2020-05-18	2020-05-18	
2001	Leonardo Warner	US	5	NaN	2020-05-31	2020-05-31	
2007	Patrice Gibso	US	5	NaN	2020-05-29	2020-05-29	
2083	Kenneth Bryan	US	5	NaN	2020-05-21	2020-05-21	
2264	ELIZABETH PAVIA	US	5	Very easy ordering and customer service was excellent. Items ordered were	2020-05-06	2020-05-06	

	name	country_origin	number_stars	review_text	date_experience	date_review	time_l
				exactly as expected.			
2040	Cherish Field	US	5	It was easy to make an order online. It received my order in good time.	2020-05-25	2020-05-25	

In [416... `df_may2020.review_text.info()`

```
<class 'pandas.core.series.Series'>
Int64Index: 283 entries, 1995 to 2283
Series name: review_text
Non-Null Count  Dtype
-----
233 non-null    object
dtypes: object(1)
memory usage: 4.4+ KB
```

In [427... `df_may2020[(df_may2020['number_stars']==5) & pd.isna(df_may2020['review_text'])]`

Out [427] :		name	country_origin	number_stars	review_text	date_experience	date_review	time_betw
	1995	Tammy FLANAGAN	US	5	NaN	2020-05-31	2020-05-31	
	2001	Leonardo Warner	US	5	NaN	2020-05-31	2020-05-31	
	2002	Jessica	US	5	NaN	2020-05-31	2020-05-31	
	2007	Patrice Gibso	US	5	NaN	2020-05-29	2020-05-29	
	2022	RAMONA KILLINGS	US	5	NaN	2020-05-27	2020-05-27	
	2026	NIKITRA GLEN	US	5	NaN	2020-05-26	2020-05-26	
	2042	Natasha Brow	US	5	NaN	2020-05-25	2020-05-25	
	2048	Shannon Allen	US	5	NaN	2020-05-24	2020-05-24	
	2058	Marquis Hine	US	5	NaN	2020-05-23	2020-05-23	
	2064	Ruth Gra	US	5	NaN	2020-05-22	2020-05-22	
	2073	Chekina Davi	US	5	NaN	2020-05-22	2020-05-22	
	2078	Marquiuis Jenkins	DE	5	NaN	2020-05-22	2020-05-22	
	2083	Kenneth Bryan	US	5	NaN	2020-05-21	2020-05-21	
	2087	Guest	US	5	NaN	2020-05-21	2020-05-21	
	2089	courtney davi	US	5	NaN	2020-05-21	2020-05-21	
	2091	BRUCE STREICH	US	5	NaN	2020-05-21	2020-05-21	
	2095	david larde	US	5	NaN	2020-05-21	2020-05-21	
	2107	Erica Morgan	US	5	NaN	2020-05-20	2020-05-20	
	2116	Suzanne Fichter	US	5	NaN	2020-05-20	2020-05-20	
	2130	Amanda	US	5	NaN	2020-05-19	2020-05-19	
	2133	Ben A. Bos	US	5	NaN	2020-05-19	2020-05-19	
	2144	customer	US	5	NaN	2020-05-18	2020-05-18	
	2146	Sarah Jones	US	5	NaN	2020-05-18	2020-05-18	
	2153	Earline Gaskins	US	5	NaN	2020-05-17	2020-05-17	
	2163	Dorene Zele	US	5	NaN	2020-05-16	2020-05-16	
	2167	aretha thomas	US	5	NaN	2020-05-16	2020-05-16	
	2172	Janet Harsh	US	5	NaN	2020-05-16	2020-05-16	
	2175	Rachel Johnso	US	5	NaN	2020-05-16	2020-05-16	

	name	country_origin	number_stars	review_text	date_experience	date_review	time_betw
2181	Yelitza Navarr	US	5	NaN	2020-05-15	2020-05-15	
2183	Shanteka Stubb	US	5	NaN	2020-05-15	2020-05-15	
2188	Will rega	US	5	NaN	2020-05-15	2020-05-15	
2193	Jason Mcge	US	5	NaN	2020-05-14	2020-05-14	
2209	Rex Perry	US	5	NaN	2020-05-13	2020-05-13	
2212	NonnieB	US	5	NaN	2020-05-13	2020-05-13	
2215	Ms. Parker	US	5	NaN	2020-05-13	2020-05-13	
2230	Sylvia Malon	US	5	NaN	2020-05-11	2020-05-11	
2240	Ultimo De Oliveira	US	5	NaN	2020-05-09	2020-05-09	
2247	Carla Gadd	US	5	NaN	2020-05-09	2020-05-09	
2257	Anna Sochacki	US	5	NaN	2020-05-07	2020-05-07	
2261	Wanda For	US	5	NaN	2020-05-06	2020-05-06	
2276	Katalin Csizmadia	US	5	NaN	2020-05-02	2020-05-02	
2279	LEE KLEI	KR	5	NaN	2020-05-01	2020-05-01	
2282	Colie Beasley	US	5	NaN	2020-05-01	2020-05-01	
2283	Daisy jane Ro	US	5	NaN	2020-05-01	2020-05-01	

50 out of 283 entries for review texts are NaNs, which is strange. Trustpilot writes on their website that it invests a lot into fighting and removing fake reviews:

<https://www.trustpilot.com/trust/combating-fake-reviews> We do not have enough data to confirm that all the 50 reviews were somehow constructed or by a bot or by a group of people with an interest to promote the online shop. This phenomenon could also be explained by the customer behavior during corona, just before Mothers' Day. This would need further investigation.

Let's compare this peak to the one in February 2017.

```
In [429... df_feb2017 = df_copy[(df_copy['month']==2) & (df_copy['year']==2017)]
```

```
In [430... df_feb2017.head()
```

Out [430] :	name	country_origin	number_stars	review_text	date_experience	date_review	time_betw_e
	4927	Tim G.	US	5	Selection and service are always great! The free 2nd day air deal is much appreciated. A few sales would be nice.	2017-02-28	2017-02-28
	4928	Diane L	US	2	I received a charm booklet with my last purchase. The charm featured was Faith, Hope, & Love \$25.00 #790119. Tried to purchase and was told it doesn't exist. Now come on.....	2017-02-28	2017-02-28
	4929	Shemika Paul	US	5	Love my product and timely delivery.	2017-02-28	2017-02-28
	4930	Alex Bar	US	5	I received my bracelet and it was the wrong size. I returned it for replacement. I received a text informing me that it had been received, I also received a text telling me that my bracelet was on its way back to me. It appears that this company is very good to there customers. I will order from them in the future.	2017-02-28	2017-02-28
	4931	Emily Nixo	US	5	Made exchanging easy.	2017-02-28	2017-02-28

In [431]: `df_feb2017.review_text.info()`

```
<class 'pandas.core.series.Series'>
Int64Index: 233 entries, 4927 to 5163
Series name: review_text
Non-Null Count  Dtype
-----
229 non-null    object
dtypes: object(1)
memory usage: 3.6+ KB
```

Here only 4 out of 233 reviews were empty, while in May 2020 50 out of 283 5-star reviews were empty. This number confirms that the reviews from May 2020 would need further analysis.

5. Summary

Most customers of the Pandora online shop seem to be very satisfied: the average rating is almost 4.7 out of 5 stars. The majority of the reviews, both positive and negative, is written on the same day as the product arrives. A closer look at the distribution of the 5-star reviews over the year reveals that there are seasons when there are more reviews written, with a big peak in December when people get gifts for the holidays. Also before Mother's Day in May, almost every year there was a small peak. In May 2020 there was an extraordinarily high number of 5-star reviews written, which is slightly suspicious because 50 out of 283 rows had empty review texts. This phenomenon would need further investigation.