Showing the first 5 rows of the dataframe as a peak into the data

	CUSTNAME	GenderCode	ADDRESS1	CITY	STATE	COUNTRY_CODE	POSTAL_CODE	POS
0	Allen Perl	Mr.	4707 Hillcrest Lane	Abeto	PG	IT	6040	
1	Allen Perl	Mr.	4707 Hillcrest Lane	Abeto	PG	IT	6040	
2	Allen Perl	Mr.	4707 Hillcrest Lane	Abeto	PG	IT	6040	
3	Allen Perl	Mr.	4707 Hillcrest Lane	Abeto	PG	IT	6040	
4	Allen Perl	Mr.	4707 Hillcrest Lane	Abeto	PG	IT	6040	

5 rows × 62 columns

Steps for data analysis checklist

1. Most number of Products ordered - Completed

2. Explore location and see where we get the highest number of orders - Completed

3. Explore date and time and find in-depth analysis of what products are bought during which frequency of time - Completed

4. Which genders tend to order what type of products - Completed

5. How long does it take for an order to arrive. Do the following factors have anything to do with it: Location, time of year, product type - Incomplete

Summary of the data

What are the dimensions of the dataframe

What are it's features

Any missing values?

Number ofnique values

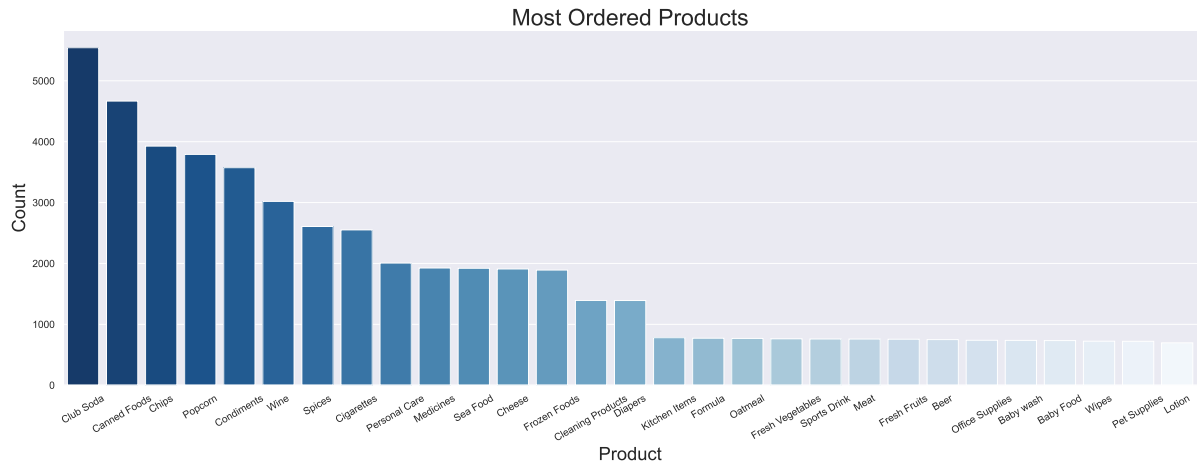
```
Rows     : 13733
Columns  : 62

Features :
     : ['CUSTNAME', 'GenderCode', 'ADDRESS1', 'CITY', 'STATE', 'COUNTRY_COD
E', 'POSTAL_CODE', 'POSTAL_CODE_PLUS4', 'ADDRESS2', 'EMAIL_ADDRESS', 'PHONE_N
UMBER', 'CREDITCARD_TYPE', 'LOCALITY', 'SALESMAN_ID', 'NATIONALITY', 'NATIONA
L_ID', 'CREDITCARD_NUMBER', 'DRIVER_LICENSE', 'CUST_ID', 'ORDER_ID', 'ORDER_D
ATE', 'ORDER_TIME', 'FREIGHT_CHARGES', 'ORDER_SALESMAN', 'ORDER_POSTED_DATE',
'ORDER_SHIP_DATE', 'AGE', 'ORDER_VALUE', 'T_TYPE', 'PURCHASE_TOUCHPOINT', 'PU
RCHASE_STATUS', 'ORDER_TYPE', 'GENERATION', 'Baby Food', 'Diapers', 'Formul
a', 'Lotion', 'Baby wash', 'Wipes', 'Fresh Fruits', 'Fresh Vegetables', 'Bee
r', 'Wine', 'Club Soda', 'Sports Drink', 'Chips', 'Popcorn', 'Oatmeal', 'Medi
cines', 'Canned Foods', 'Cigarettes', 'Cheese', 'Cleaning Products', 'Condime
nts', 'Frozen Foods', 'Kitchen Items', 'Meat', 'Office Supplies', 'Personal C
are', 'Pet Supplies', 'Sea Food', 'Spices']

Missing values    : 43702

Unique values :
 CUSTNAME           4698
GenderCode             4
ADDRESS1            4709
CITY                2197
STATE                145
                    ...
Office Supplies        2
Personal Care          2
Pet Supplies           2
Sea Food               2
Spices                 2
Length: 62, dtype: int64

Missing values in columns:
 CUSTNAME               0
GenderCode              0
ADDRESS1                0
CITY                    0
STATE                1587
                    ...
Office Supplies         0
Personal Care           0
Pet Supplies            0
Sea Food                0
Spices                  0
Length: 62, dtype: int64

Duplicated rows:
 False    13733
dtype: int64
```

```
<ipython-input-5-7baa07d75a02>:9: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy
  df1['GenderCode'][i] = 'M';
<ipython-input-5-7baa07d75a02>:6: SettingWithCopyWarning:
A value is trying to be set on a copy of a slice from a DataFrame

See the caveats in the documentation: https://pandas.pydata.org/pandas-docs/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy
  df1['GenderCode'][i] = 'F';
```

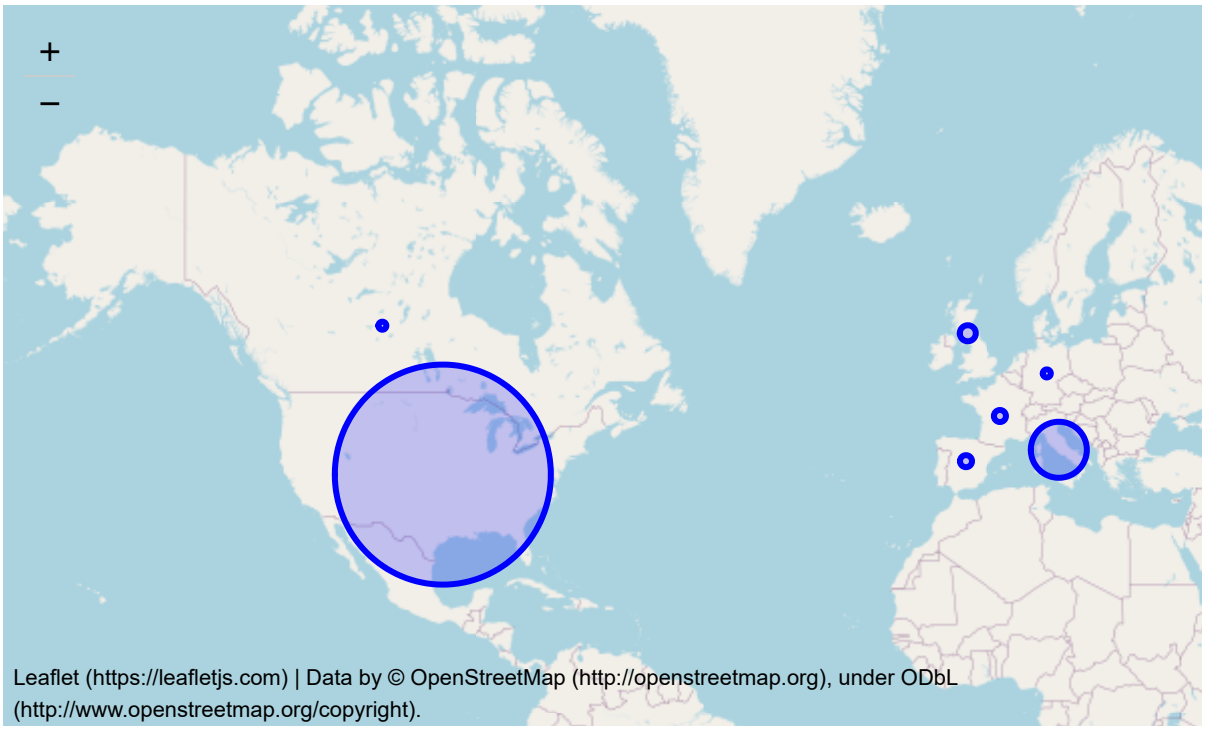
	GenderCode	CITY	STATE	COUNTRY_CODE	ORDER_ID	ORDER_DATE	ORDER_TI
0	F	Alcalali	NaN	ES	4950	25-09-16 0:00	25-09 16
1	M	Aldearrubia	NaN	ES	3363	25-11-16 0:00	25-11 16
2	F	Angas Plains	SA	AU	4942	16-03-16 17:47	16-03 17
3	M	Antioch	WI	US	479	28-03-16 0:00	28-03-16 2
4	M	Antioch	WI	US	2971	29-02-16 0:00	29-02 22
...	...	...	...	...	...	...	
52557	F	Los Angeles	CA	US	15956	13-02-16 0:00	13-02 19
52558	M	Los Angeles	CA	US	15973	05-07-16 0:00	05-07-16 8
52559	F	Los Angeles	CA	US	15980	19-11-16 0:00	19-11 13
52560	F	Los Angeles	CA	US	15984	28-04-16 0:00	28-04 20
52561	F	Los Angeles	CA	US	15996	20-11-16 0:00	20-11-16 4

52562 rows × 18 columns

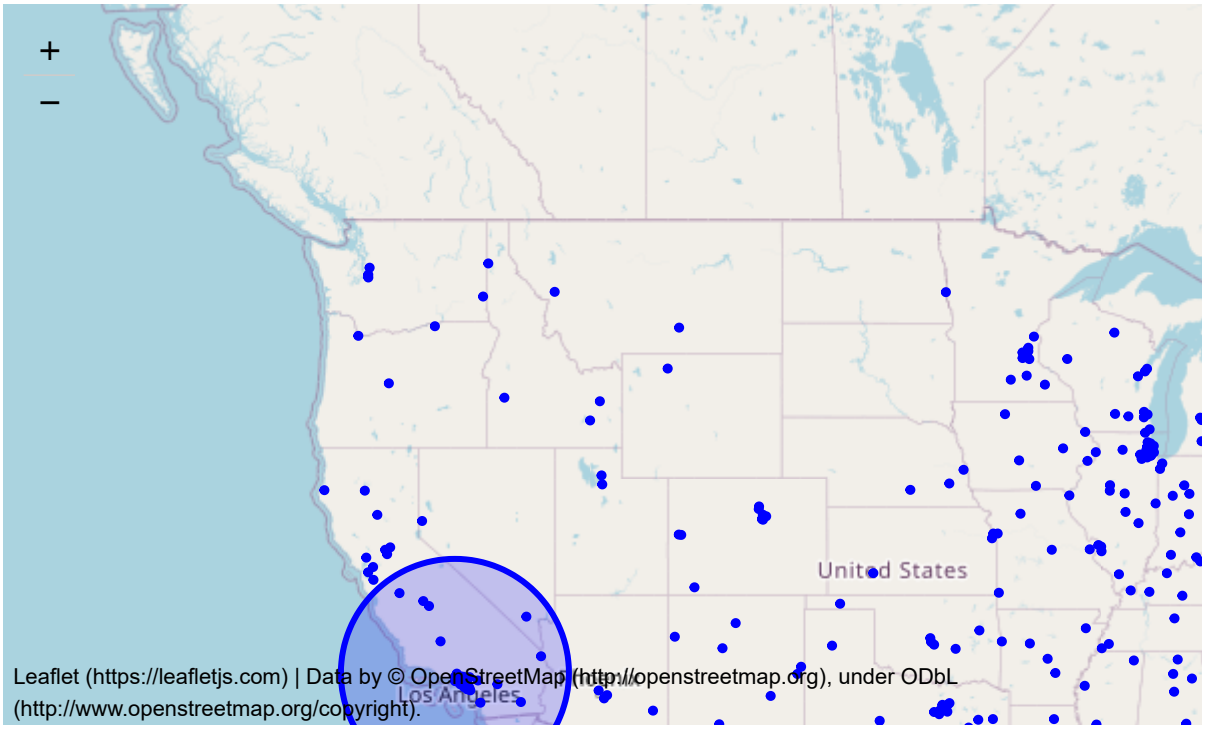
1. What are the Most Ordered Products



2. Where are These Orders Coming From?



The US seems to be our largest market. Lets take a deeper dive into the US and further narrow down the location.



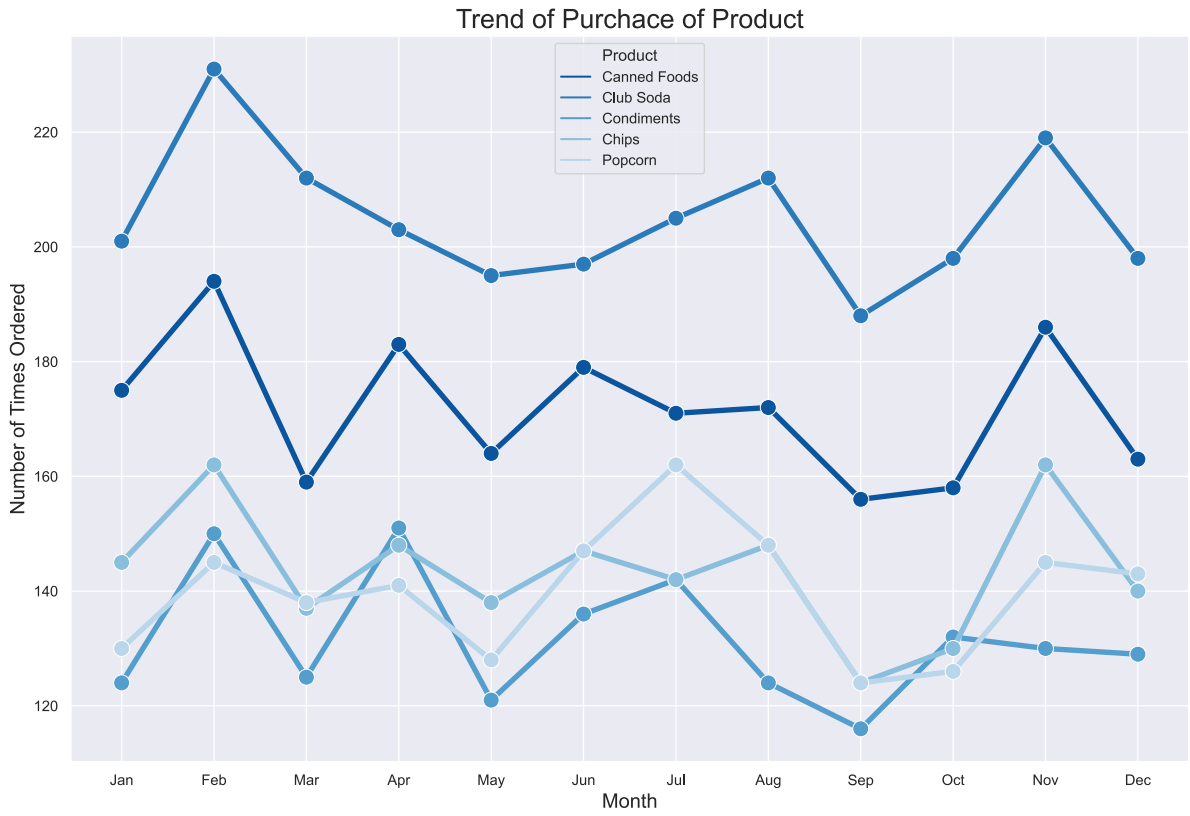
Here we see that Los Angeles comes out on top more than any other city by a large margin

Within L.A which products are ordered the most - Top 5

3. What do the sale of these products look like over time?

```
<ipython-input-15-ae2c5d47c216>:5: 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/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy
  la['ORDER_DATE'] = la['ORDER_DATE'].str[:8];
<ipython-input-15-ae2c5d47c216>:8: 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/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy
  la['ORDER_DATE'] = pd.to_datetime(la['ORDER_DATE'], format = '%d-%m-%y');
<ipython-input-15-ae2c5d47c216>:9: 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/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy
  la['ORDER_TIME'] = pd.to_datetime(la['ORDER_TIME'], format = '%d-%m-%y %H:%
M');
<ipython-input-15-ae2c5d47c216>:12: 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/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy
  la['Year'] = la['ORDER_DATE'].dt.year;
<ipython-input-15-ae2c5d47c216>:13: 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/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy
  la['Month'] = la['ORDER_DATE'].dt.month;
<ipython-input-15-ae2c5d47c216>:14: 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/s
table/user_guide/indexing.html#returning-a-view-versus-a-copy
  la['Month Name'] = la['ORDER_DATE'].dt.strftime('%b');
```

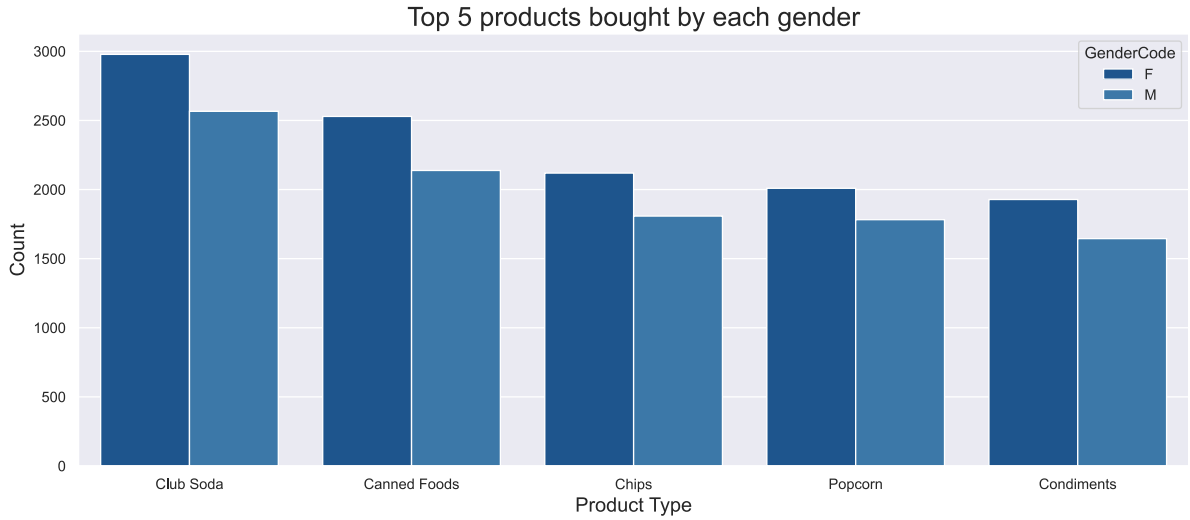
```
Text(0, 0.5, 'Number of Times Ordered')
```



This figure can give stores a good representation of what products to stock up on before the demand comes in

4. Which genders tend to order which type of products


Products ordered by Gender
25000
20000
15000
10000
5000
0
Count
Male
Female
Gender


Top 5 products bought by each gender
GenderCode
F
M
3000
2500
2000
1500
1000
500
0
Count
Club Soda
Canned Foods
Chips
Popcorn
Condiments
Product Type

From the second figure we can see that women tend to order products online more than men

The top 5 categories for each are plotted in the second figure and women have consistently ordered more than men in each category