

# 1. Who is The Market Leader?

In [1]:

```
# Import library
import numpy as np
import pandas as pd
import matplotlib.pyplot as plt
```

In [2]:

```
# Load data
df = pd.read_csv("raw data.csv")
```

In [3]:

```
# Melihat 5 data teratas
df.head()
```

Out[3]:

Unnamed: 0		nama	harga	seller	terjual	tanggal_ambil_data
0	1	Bertolli Extra Virgin Olive Oil / Minyak Zaitu...	Rp85.000	Food Republic	Terjual 1.779 Produk	2020-09-04
1	2	Original Extra Virgin Olive Oil (EVOO) Casa Di...	Rp60.000	Yubyre	Terjual 1.624 Produk	2020-09-04
2	3	EVOO Baby Olive Oil Casa Di Oliva Olivia For K...	Rp58.000	Papamama Babyshop	Terjual 2.161 Produk	2020-09-04
3	4	Casa di Oliva - Extra Virgin Olive Oil for Kid...	Rp59.000	Chubby Baby Shop	Terjual 1.696 Produk	2020-09-04
4	5	BEST PRICE CASA DI OLIVA EXTRA VIRGIN OLIVE OI...	Rp59.900	Whiz World	Terjual 3.115 Produk	2020-09-04

In [4]:

```
# Mengecek keberadaan null
df.info()
```

```
<class 'pandas.core.frame.DataFrame'>
RangeIndex: 3259 entries, 0 to 3258
Data columns (total 6 columns):
#   Column                Non-Null Count  Dtype
---  -
0   Unnamed: 0            3259 non-null   int64
1   nama                  3259 non-null   object
2   harga                 3259 non-null   object
3   seller                3259 non-null   object
4   terjual               3259 non-null   object
5   tanggal_ambil_data    3259 non-null   object
dtypes: int64(1), object(5)
memory usage: 89.2+ KB
```

In [5]:

```
# Membuang kolom "Unnamed: 0"
df = df.iloc[:, 1:6]

# Mengedit nama column
df.columns = ["Nama Produk", "Harga", "Seller", "Terjual", "Tanggal Ambil Data"]
```

In [6]:

```
# Membuat array untuk menyimpan data dari column "Harga" dan "Terjual" dalam format integer
harga_int = [0 for i in range(len(df))]
terjual_int = [0 for i in range(len(df))]

for i in range(len(df)):
    # Memfilter agar hanya angka yang terdapat di setiap elemen array
    harga_int[i] = "".join(x for x in df["Harga"][i] if x.isdigit())
    terjual_int[i] = "".join(x for x in df["Terjual"][i] if x.isdigit())

    # Mengubah tipe data dari string menjadi integer
    harga_int[i] = int(harga_int[i])
    terjual_int[i] = int(terjual_int[i])
```

In [7]:

```
# Recheck
print(harga_int[:5], harga_int[-5:]) # Mengecek 5 data pertama harga_int
print(terjual_int[:5], terjual_int[-5:]) # Mengecek 5 data pertama terjual_int
```

```
[85000, 60000, 58000, 59000, 59900] [70000, 65000, 185000, 67500, 58000]
[1779, 1624, 2161, 1696, 3115] [34, 24, 20, 15, 18]
```

In [8]:

```
# Mengganti setiap nilai pada column "Harga" dan "Terjual" dengan int dari hasil looping sebelumnya
for i in range(len(df)):
    df["Harga"][i] = harga_int[i]
    df["Terjual"][i] = terjual_int[i]

# Recheck
df.head()
```

Out[8]:

	Nama Produk	Harga	Seller	Terjual	Tanggal Ambil Data
0	Bertolli Extra Virgin Olive Oil / Minyak Zaitu...	85000	Food Republic	1779	2020-09-04
1	Original Extra Virgin Olive Oil (EVOO) Casa Di...	60000	Yubyre	1624	2020-09-04
2	EVOO Baby Olive Oil Casa Di Oliva Olivia For K...	58000	Papamama Babysshop	2161	2020-09-04
3	Casa di Oliva - Extra Virgin Olive Oil for Kid...	59000	Chubby Baby Shop	1696	2020-09-04
4	BEST PRICE CASA DI OLIVA EXTRA VIRGIN OLIVE OI...	59900	Whiz World	3115	2020-09-04

In [9]:

```
# List Nama Minyak Zaitun
merk = ["Le Riche", "Bertolli", "Borges", "Ondoliva Selection", "Gallo",
        "Rafael Salgado", "Olitalia", "Bariani White Truffle", "Pompeian",
        "Filippo Berio", "Mueloliva", "Sasso", "Merula", "RS Olive Oil Pomace",
        "La Tourangelle", "Casa di Oliva", "Casa di Olivia", "HPA", "Medina",
        "Cobram Estate", "Pietro Coricelli", "Yummy Bites", "Orkide", "Arbequina",
        "Pomace"]

# Deklarasi array baru tempat menyimpan nama produk
product_merk = [0 for i in range(len(df))]

# Mengecek satu per satu nama dari masing-masing produk di column "Nama Produk"
# dengan masing-masing elemen di array "merk"
for i in range(len(df)):
    for j in range(len(merk)):
        if (merk[j].replace(" ", "").lower() in df["Nama Produk"][i].replace(" ", "").lower()):
            product_merk[i] = merk[j]

    if product_merk[i] == 0:
        product_merk[i] = "others"

# Menambahkan column "Merk" di sebelah column "Nama Produk"
```

```
df.insert(1, "Merk", product_merk, True)
```

```
# Recheck
df.head()
```

Out[9]:

	Nama Produk	Merk	Harga	Seller	Terjual	Tanggal Ambil Data
0	Bertolli Extra Virgin Olive Oil / Minyak Zaitu...	Bertolli	85000	Food Republic	1779	2020-09-04
1	Original Extra Virgin Olive Oil (EVOO) Casa Di...	Casa di Olivia	60000	Yubyre	1624	2020-09-04
2	EVOO Baby Olive Oil Casa Di Oliva Olivia For K...	Casa di Oliva	58000	Papamama Babyshop	2161	2020-09-04
3	Casa di Oliva - Extra Virgin Olive Oil for Kid...	Casa di Oliva	59000	Chubby Baby Shop	1696	2020-09-04
4	BEST PRICE CASA DI OLIVA EXTRA VIRGIN OLIVE OL...	Casa di Oliva	59900	Whiz World	3115	2020-09-04

In [10]:

```
# Karena "Casa di Olivia" dan "Casa di Oliva" adalah produk yang sama, maka akan
# dibuat menjadi satu merk yang sama
for i in range(len(df)):
    if df["Merk"][i] == "Casa di Olivia":
        df["Merk"][i] = "Casa di Oliva"

# Recheck
df.head()
```

Out[10]:

	Nama Produk	Merk	Harga	Seller	Terjual	Tanggal Ambil Data
0	Bertolli Extra Virgin Olive Oil / Minyak Zaitu...	Bertolli	85000	Food Republic	1779	2020-09-04
1	Original Extra Virgin Olive Oil (EVOO) Casa Di...	Casa di Oliva	60000	Yubyre	1624	2020-09-04
2	EVOO Baby Olive Oil Casa Di Oliva Olivia For K...	Casa di Oliva	58000	Papamama Babyshop	2161	2020-09-04
3	Casa di Oliva - Extra Virgin Olive Oil for Kid...	Casa di Oliva	59000	Chubby Baby Shop	1696	2020-09-04
4	BEST PRICE CASA DI OLIVA EXTRA VIRGIN OLIVE OL...	Casa di Oliva	59900	Whiz World	3115	2020-09-04

In [11]:

```
# Deklarasi array untuk menyimpang total penjualan dari masing-masing produk
product = []
sales = []

# Meletakkan setiap merk ke array "product" dan hasil penjualan yang bersangkutan ke arra
y "sales"
for i in range(len(df)):
    if df["Merk"][i] not in product:
        product += [df["Merk"][i]]
        sales += [0]

    for j in range(len(product)):
        if product[j] == df["Merk"][i]:
            sales[j] = sales[j] + (df["Harga"][i] * df["Terjual"][i])

# Mengurutkan secara descending
for i in range(len(sales)):
    for j in range(len(sales)):
        if sales[i] > sales[j]:
```

```
sales[i], sales[j] = sales[j], sales[i]
product[i], product[j] = product[j], product[i]
```

In [12]:

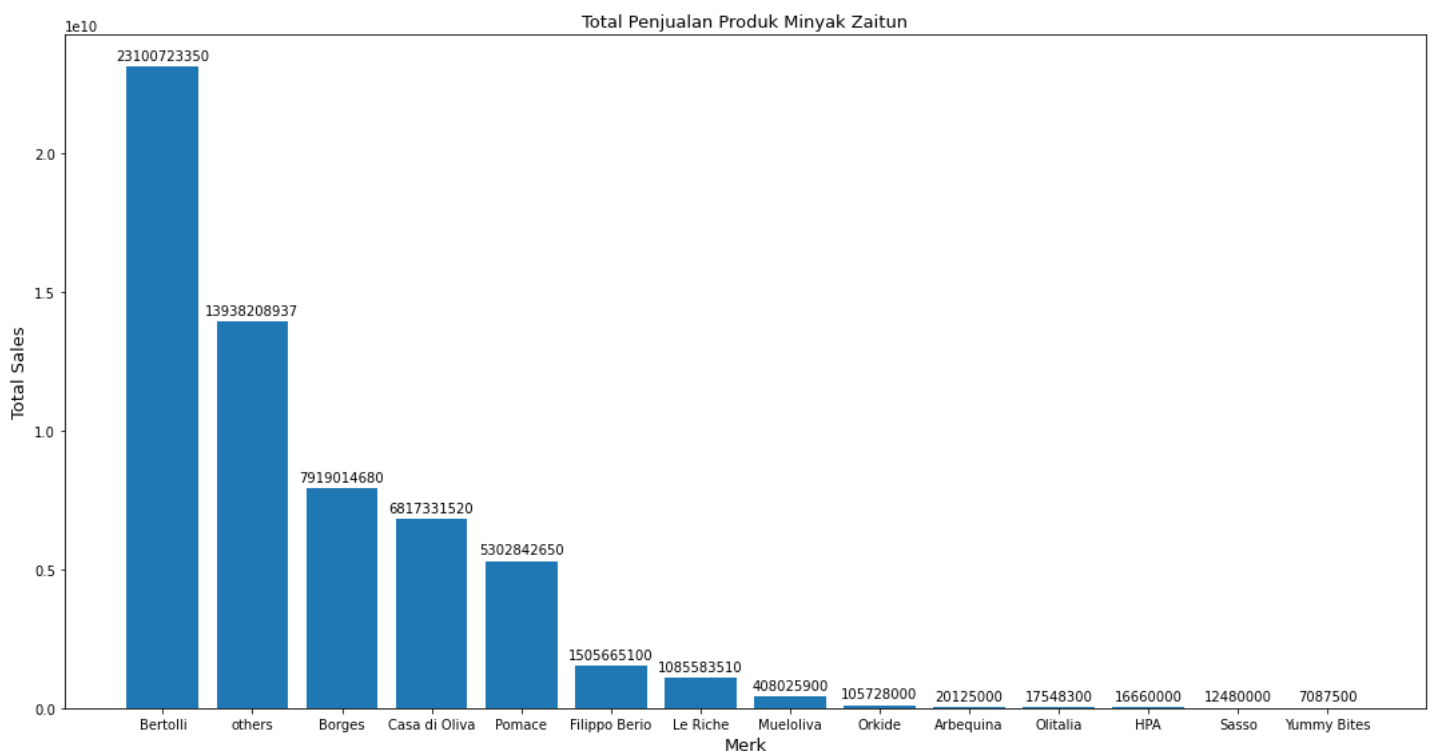
```
fig, ax = plt.subplots(figsize=(15, 8))

bar1 = ax.bar(product, sales)
ax.set_title("Total Penjualan Produk Minyak Zaitun", size=13)
ax.set_xlabel("Merk", size=13)
ax.set_ylabel("Total Sales", size=13)

# Untuk menambahkan label di atas barplot lihat https://matplotlib.org/3.1.1/gallery/line_s_bars_and_markers/barchart.html
def autolabel(rects):
    """Attach a text label above each bar in *rects*, displaying its height."""
    for rect in rects:
        height = rect.get_height()
        ax.annotate('{}'.format(round(height, 2)),
                    xy=(rect.get_x() + rect.get_width() / 2, height),
                    xytext=(0, 3), # 3 points vertical offset
                    textcoords="offset points",
                    ha='center', va='bottom')

autolabel(bar1)
fig.tight_layout()

plt.show()
fig.savefig("Penjualan Minyak Zaitun")
```



- Setelah melalui beberapa proses cleaning dan perhitungan, dapat disimpulkan bahwa brand dengan sales value terbesar dalam rentang waktu tersebut adalah Bertolli dengan total pendapatan mencapai **Rp23.100.723.350**

In [13]:

```
# Mengganti tipe data pada column "Harga" dan "Terjual" dari object to float
df[["Harga", "Terjual"]] = df[["Harga", "Terjual"]].astype("float")

# Mengambil 5 nama produk dengan penjualan tertinggi
big_five = product[:5]

# Menyimpang rata-rata harga dan total penjualan kelima brand dan dikelompokkan berdasarkan "Tanggal Ambil Data"
group = [0 for i in range(5)]
```

```
for i in range(len(big_five)):
    temp = df[df["Merk"] == big_five[i]].groupby("Tanggal Ambil Data").agg({"Harga": "mean", "Terjual": "sum"})
    group[i] = [temp["Harga"], temp["Terjual"]]
```

In [14]:

```
# Menghitung total penjualan pada kelima brand untuk masing-masing tanggal dan disimpan ke dalam array y1 - y5
a = len(df.groupby("Tanggal Ambil Data").sum())
x = np.arange(a)
y = [0 for i in range(len(group))]
for i in range(len(group)):
    y[i] = [group[i][0][j] * group[i][1][j] for j in range(len(group[i][0]))]

fig, ax = plt.subplots(figsize=(12.3, 6.5))
for i in range(len(group)):
    ax.bar(x+(i*0.1), y[i], width=0.1, label=big_five[i])

ax.set_xticks(x+0.2)
ax.set_xticklabels(df.groupby("Tanggal Ambil Data").sum().index[:])
ax.set_ylabel("Total Sales")
ax.set_title("Penjualan Produk Minyak Zaitun per Tanggal")

fig.tight_layout()

fig.legend(bbox_to_anchor=(1.115, 1.085))

plt.show()
fig.savefig("Penjualan Minyak Zaitun per tanggal")
```



- Dari 5 brand dengan total sales tertinggi, dilihat secara terpisah per tanggal yang tersedia pada dataframe bahwa total penjualan untuk **Bertolli** selalu memimpin dari waktu ke waktu dalam rentang waktu tersebut (asumsi harga yang digunakan harga rata-rata)
- Maka, dapat disimpulkan bahwa Market Leader dalam rentang waktu tersebut adalah **### Bertolli**

## 2. Brands with the Biggest Sales Value at Big Sale 9.9

In [15]:

```
# Mengambil semua data yang diambil pada tanggal 2020-09-09
```

```
# Mengambil semua data yang diambil pada tanggal 2020-09-09
df_99 = df[df["Tanggal Ambil Data"] == "2020-09-09"]

# Memperbaiki urutan index
df_99.index = [i for i in range(len(df_99))]

# Recheck
df_99.head()
```

Out[15]:

	Nama Produk	Merk	Harga	Seller	Terjual	Tanggal Ambil Data
0	Bertolli Extra Virgin Olive Oil / Minyak Zaitu...	Bertolli	85000.0	Food Republic	1782.0	2020-09-09
1	Original Extra Virgin Olive Oil (EVOO) Casa Di...	Casa di Oliva	60000.0	Yubyre	1624.0	2020-09-09
2	EVOO Baby Olive Oil Casa Di Oliva Olivia For K...	Casa di Oliva	58000.0	Papamama Babysshop	2129.0	2020-09-09
3	Casa di Oliva - Extra Virgin Olive Oil for Kid...	Casa di Oliva	59000.0	Chubby Baby Shop	3.0	2020-09-09
4	BEST PRICE CASA DI OLIVA EXTRA VIRGIN OLIVE OL...	Casa di Oliva	59900.0	Whiz World	3116.0	2020-09-09

In [16]:

```
# Deklarasi array tempat menyimpan "Merk" beserta total sales yang terkait
# pada array "product_99" dan "sales_99"
product_99 = []
sales_99 = []

for i in range(len(df_99)):
    if df_99["Merk"][i] not in product_99:
        product_99 += [df_99["Merk"][i]]
        sales_99 += [0]

    for j in range(len(product_99)):
        if product_99[j] == df_99["Merk"][i]:
            sales_99[j] = sales_99[j] + (df_99["Harga"][i] * df_99["Terjual"][i])

# Mengurutkan secara descending
for i in range(len(sales_99)):
    for j in range(len(sales_99)):
        if sales_99[i] > sales_99[j]:
            sales_99[i], sales_99[j] = sales_99[j], sales_99[i]
            product_99[i], product_99[j] = product_99[j], product_99[i]
```

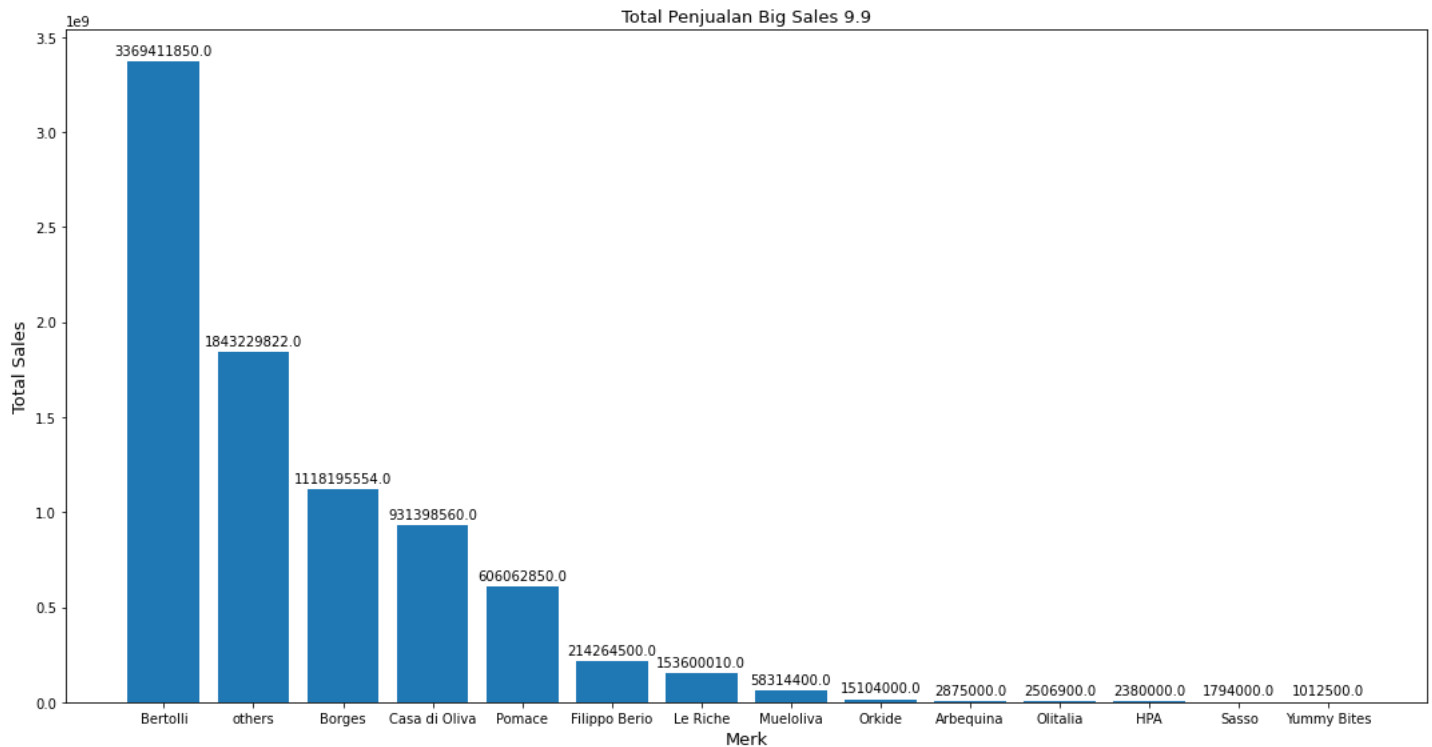
In [17]:

```
fig, ax = plt.subplots(figsize=(15, 8))
bar2 = ax.bar(product_99, sales_99)
ax.set_title("Total Penjualan Big Sales 9.9", size=13)
ax.set_xlabel("Merk", size=13)
ax.set_ylabel("Total Sales", size=13)

# Untuk menambahkan label di atas barplot lihat https://matplotlib.org/3.1.1/gallery/line_s_bars_and_markers/barchart.html
def autolabel(rects):
    """Attach a text label above each bar in *rects*, displaying its height."""
    for rect in rects:
        height = rect.get_height()
        ax.annotate('{}' .format(round(height, 2)),
                    xy=(rect.get_x() + rect.get_width() / 2, height),
                    xytext=(0, 3), # 3 points vertical offset
                    textcoords="offset points",
                    ha='center', va='bottom')

autolabel(bar2)
fig.tight_layout()
```

```
plt.show()
fig.savefig("Penjualan Minyak Zaitun 9-9")
```



- Setelah melalui beberapa proses perhitungan, dapat disimpulkan bahwa Brand dengan Sales Value terbesar selama Big Sales 9.9 adalah Bertolli dengan total pendapatan mencapai **Rp3.369.411.850**
- Jadi, jawabannya adalah **### Bertolli**

### 3. Imagine that you work in one of these oil brands, from this data what can you suggest to the sales / marketing team?

- Asumsi saya bekerja di **Bertolli**
- Secara garis besar terlihat bahwa penjualan dari brand ini sudah baik dan kian memimpin dari waktu ke waktu dalam rentang waktu tersebut
- Perkembangannya pun cenderung stabil bahkan beberapa kali mengalami peningkatan
- Jika dibandingkan dengan brand lain, Bertolli mungkin terlalu standout. Namun, kita bisa lihat bahwa gabungan dari brand-brand dengan merek yang mungkin kurang terkenal ternyata menempati posisi kedua setelah Bertolli.
- Bahkan, total sales dari waktu ke waktu dalam rentang waktu tersebut hampir menyamai Bertolli dan cenderung meningkat.
- Saya menyarankan untuk mencoba melakukan research terhadap hal-hal apa saja yang menjadi keunikan dari brand Bertolli yang tidak ada di brand lain sehingga market pasar tetap berpusat ke Bertolli
- Tidak menutup kemungkinan salah satu dari brand di dalam kelompok others ini suatu saat bisa increase massively dan menyaingi brand Bertolli.
- Salah satu sarana yang bisa dilakukan adalah mengadakan promo beberapa pembelian Bertolli dengan syarat tertentu dimana promo ini tetap menguntungkan Bertolli tetapi dengan tujuan untuk meningkatkan kuantitas penjualan.
- Sehingga jumlah customer meningkat --> sales meningkat --> Bertolli tetap memimpin market pasar