Oktoberfest EDA with Plotly


```
In [1]:  import polars as pl

         import plotly.express as px
         import plotly.offline as po
         po.init_notebook_mode()
```

```
In [2]:  df = pl.read_csv('oktoberfest.csv')
         df.head()
```

Out[2]: shape: (5, 8)

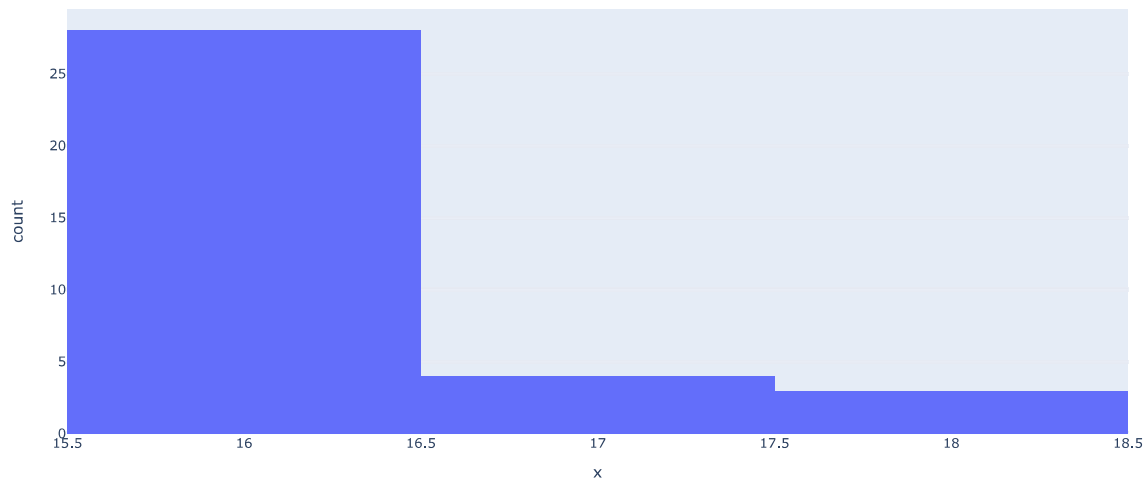| year	duration	guests_total	guests_daily	beer_price	beer_consumption	roast_chicken_price	roast_chicken_consumption
i64	i64	f64	i64	f64	i64	f64	i64
1985	16	7.1	444	3.2	54541	4.77	629520
1986	16	6.7	419	3.3	53807	3.92	698137
1987	16	6.5	406	3.37	51842	3.98	732859
1988	16	5.7	356	3.45	50951	4.19	720139
1989	16	6.2	388	3.6	51241	4.22	775674

```
In [3]:  df.describe()
```

Out[3]: shape: (7, 9)

| describe	year	duration	guests_total	guests_daily	beer_price	beer_consumption	roast_chicken_price	roast_chicken_consumption
str	f64	f64	f64	f64	f64	f64	f64	f64
"count"	35.0	35.0	35.0	35.0	35.0	35.0	35.0	35.0
"null_count"	0.0	0.0	0.0	0.0	0.0	0.0	0.0	0.0
"mean"	2002.0	16.285714	6.314286	388.171429	6.870571	62223.371429	7.592	571920.714286
"std"	10.246951	0.621735	0.390432	26.337723	2.581078	10229.542414	2.621272	123620.418372
"min"	1985.0	16.0	5.5	329.0	3.2	48698.0	3.92	351705.0
"max"	2019.0	18.0	7.1	444.0	11.71	79225.0	12.15	807710.0
"median"	2002.0	16.0	6.4	394.0	6.75	61163.0	8.12	521872.0

Days by Year

```
In [4]:  fig = px.histogram(x=df['duration'])
         po.iplot(fig, 'duration')
```
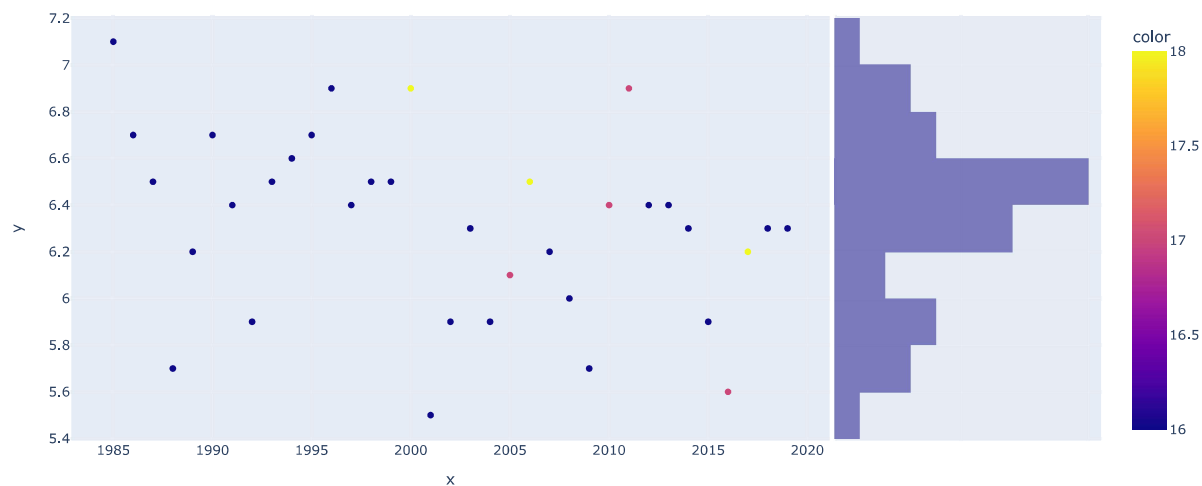
Daily Guests

```
In [5]:  fig = px.scatter(x=df['year'], y=df['guests_daily'], color=df['duration'], marginal_y='histogram')
         po.iplot(fig, 'daily-guest-vs-year')
```
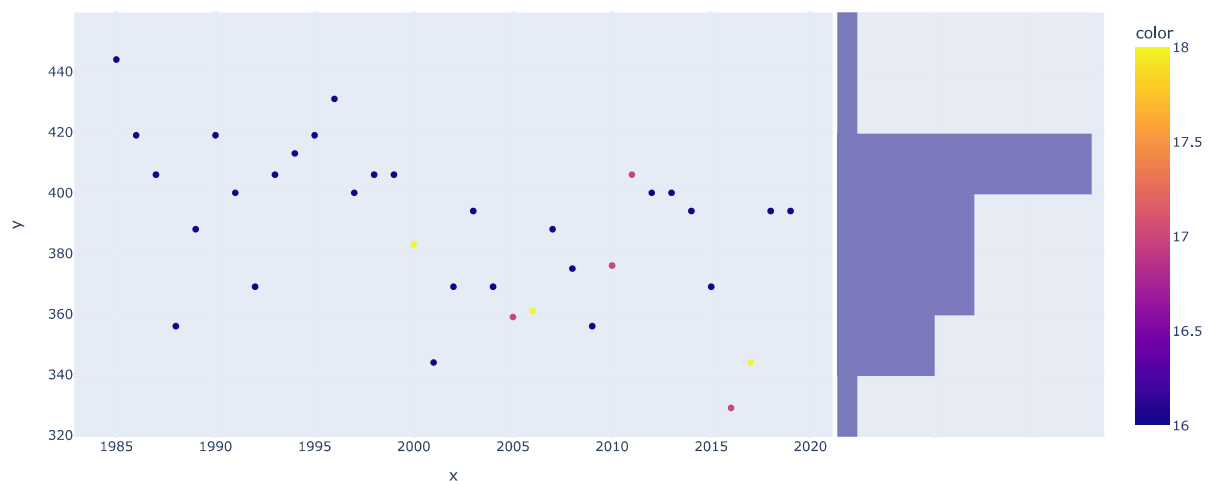
Total Guests
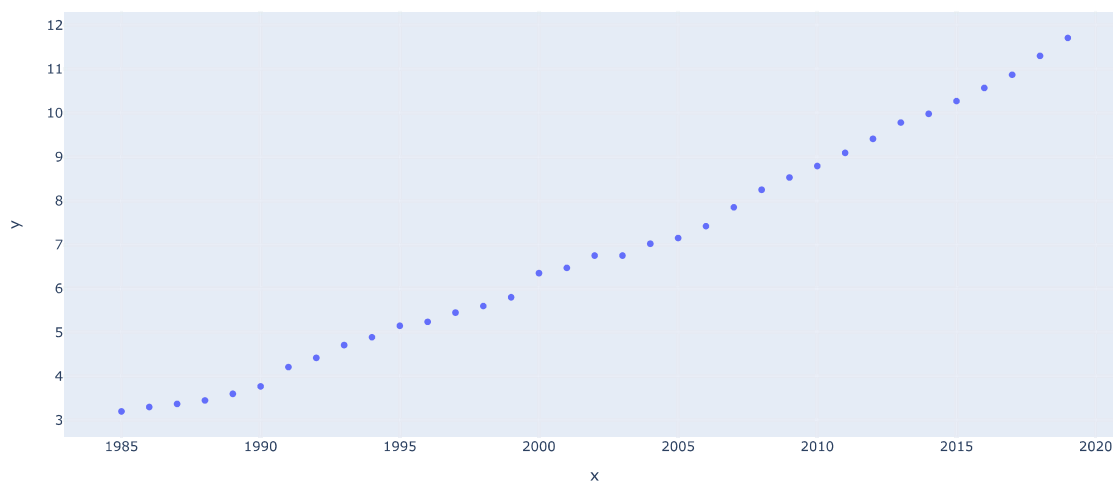
```
In [6]:  fig = px.scatter(x=df['year'], y=df['guests_total'], color=df['duration'], marginal_y='histogram')
         po.iplot(fig, 'total-guest-year')
```
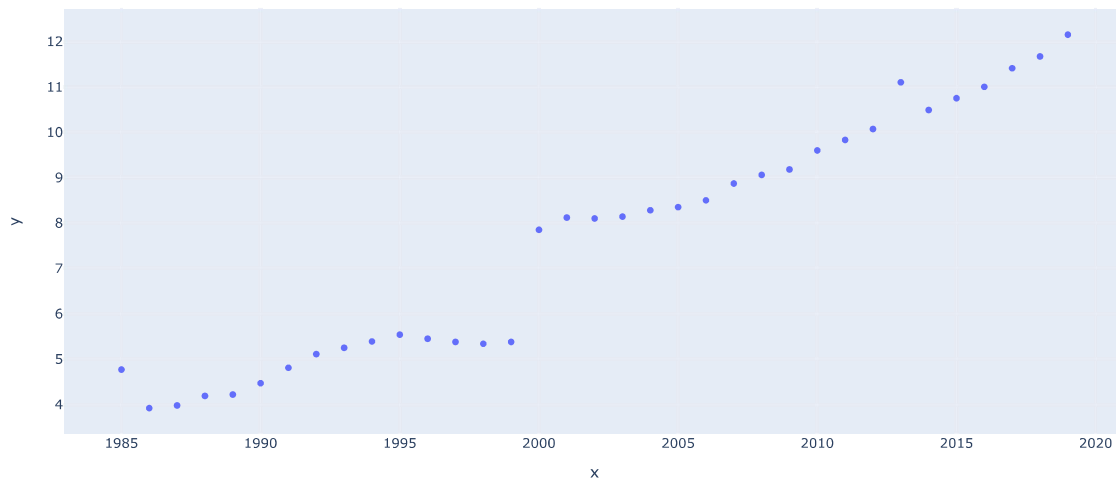
Beer Price over the Years

```
In [7]: fig = px.scatter(x=df['year'], y=df['beer_price'])
        po.iplot(fig, 'year-vs-beer-price')
```

Chicken Price over the Years
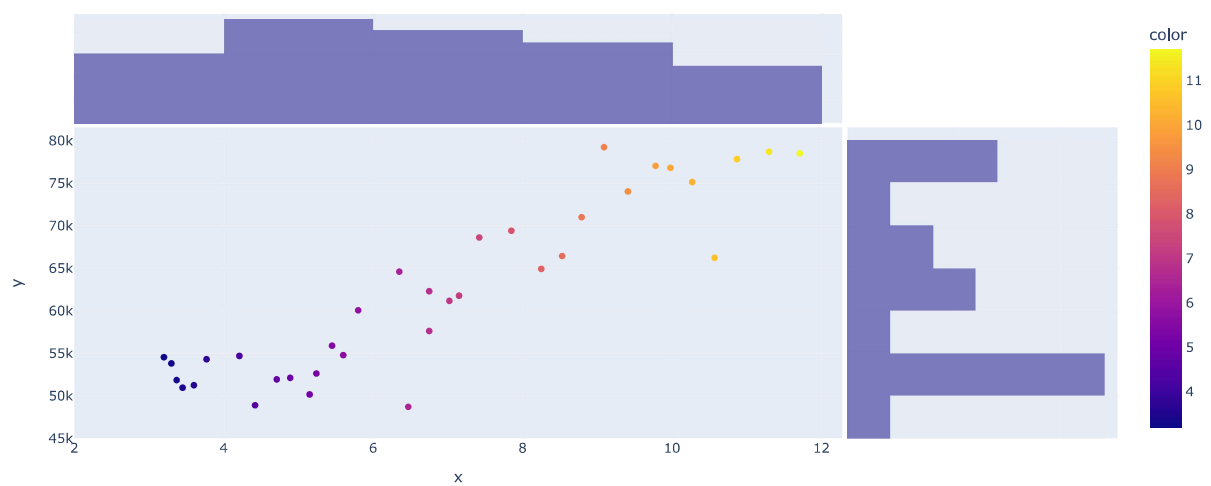
```
In [8]: fig = px.scatter(x=df['year'], y=df['roast_chicken_price'])
        po.iplot(fig, 'year-vs-roast-chicken-price')
```

Beer Price and Consumption
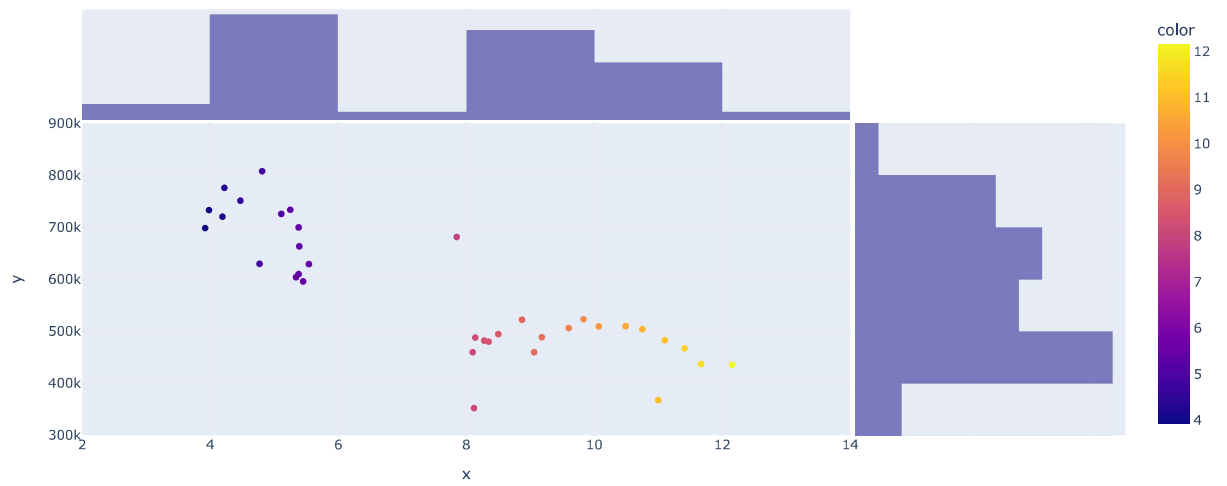
```
In [9]: fig = px.scatter(x=df['beer_price'], y=df['beer_consumption'], color=df['beer_price'],
                         marginal_x='histogram', marginal_y='histogram')
        po.iplot(fig, 'beer-price-vs-consumption')
```

Chicken Price and Consumption

```
In [10]: fig = px.scatter(x=df['roast_chicken_price'], y=df['roast_chicken_consumption'], color=df['roast_chicken_price'],
                          marginal_x='histogram', marginal_y='histogram')
         po.iplot(fig, 'chicken-price-vs-consumption')
```

In [ ]: