

Sky Econometrician Data Task.

- Mock Data on Subscription Sales and Prices

Enoch Osei - 27/08/2023

Outline.

- Introduction: What is this Data?
- Data Preparation.
- Exploratory Data Analysis (EDA).
- Package Sales Analysis.
- Region Sales Analysis.
- Temporal Analysis.
- Price Analysis.
- Commercial Insights & Actionable Takeaways.

Introduction: What is this Data?

- We have two (2) sheets;
 - “**Data**” - Containing data on the weekly sales of the various **Sky packages** and the **UK region** to which the sales belong.
 - **The Sky packages** are:
 - Cinema
 - Sports
 - Cinema & Sports Combined / Sky Premium
 - “**Price**” - Contains the weekly prices of the various Sky packages.
 - Time period: **2016-01-01** to **2020-01-24**

Sales Data Sheet

Week Commencing Friday	Package	Region	Sales
2016-01-01	Cinema	East	542
2016-01-01	Cinema & Sports	Scotland	268
2016-01-01	Cinema & Sports	South West	108
2016-01-01	Cinema	South West	385
2016-01-01	Sports	South West	210
2016-01-01	Cinema & Sports	South East	230
2016-01-01	Cinema	South East	771
2016-01-01	Sports	South East	426

Price Data Sheet

Unnamed: 0	Cinema	Sports	Combined
2016-01-01	10	20	25
2016-01-08	10	20	25
2016-01-15	10	20	25
2016-01-22	10	20	25
2016-01-29	10	20	25
2016-02-05	10	20	25
2016-02-12	10	20	25
2016-02-19	10	20	25
2016-02-26	10	20	25

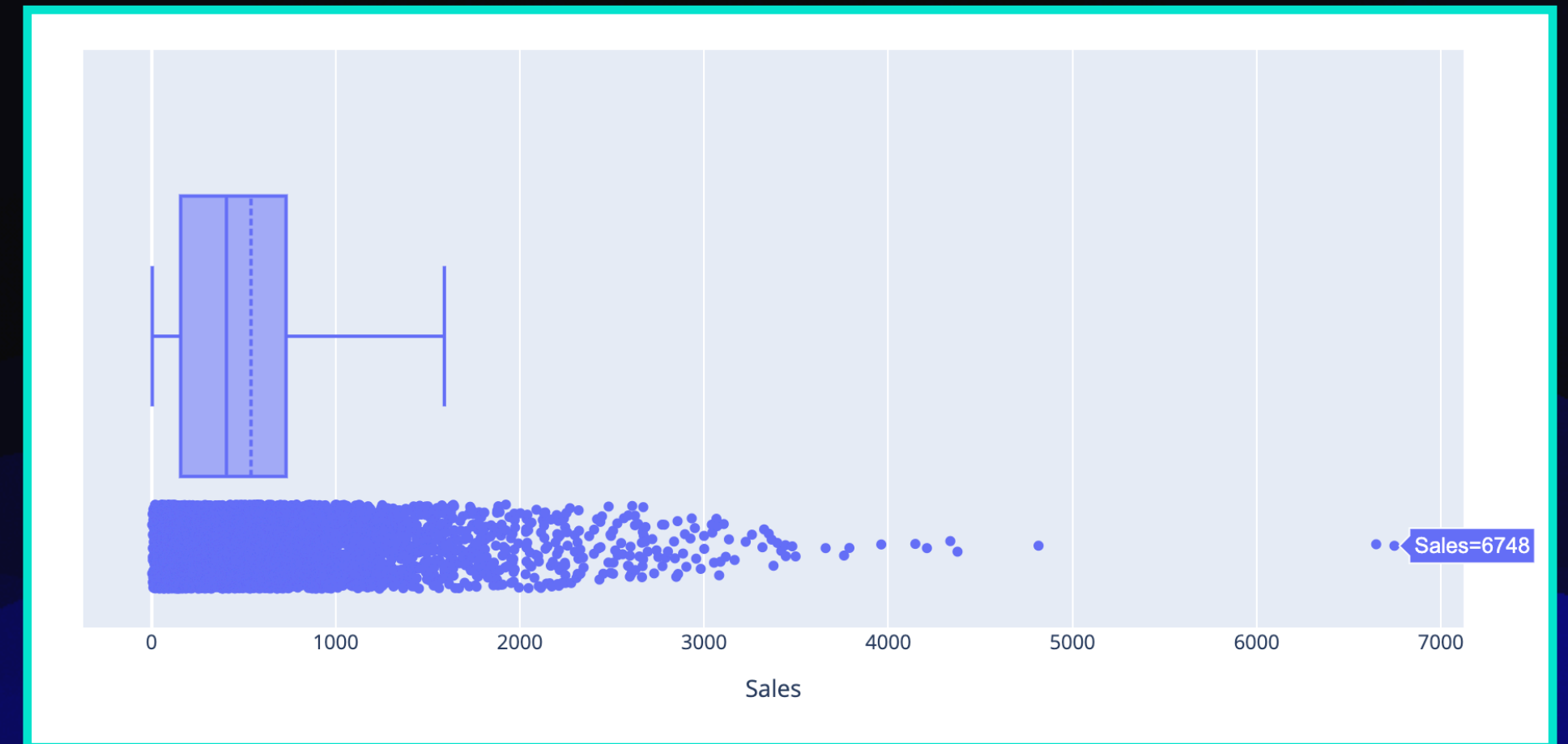
Data Preparation.

- Data was *cleaned* by:
 - Handling missing/null values.
 - Checking column data types.
 - Check consistency in categorical data for the "Data" sheet.
 - Identifying anomalies and outlier values.
- Additional “**Revenue**” column created in Sales Data from using price data.

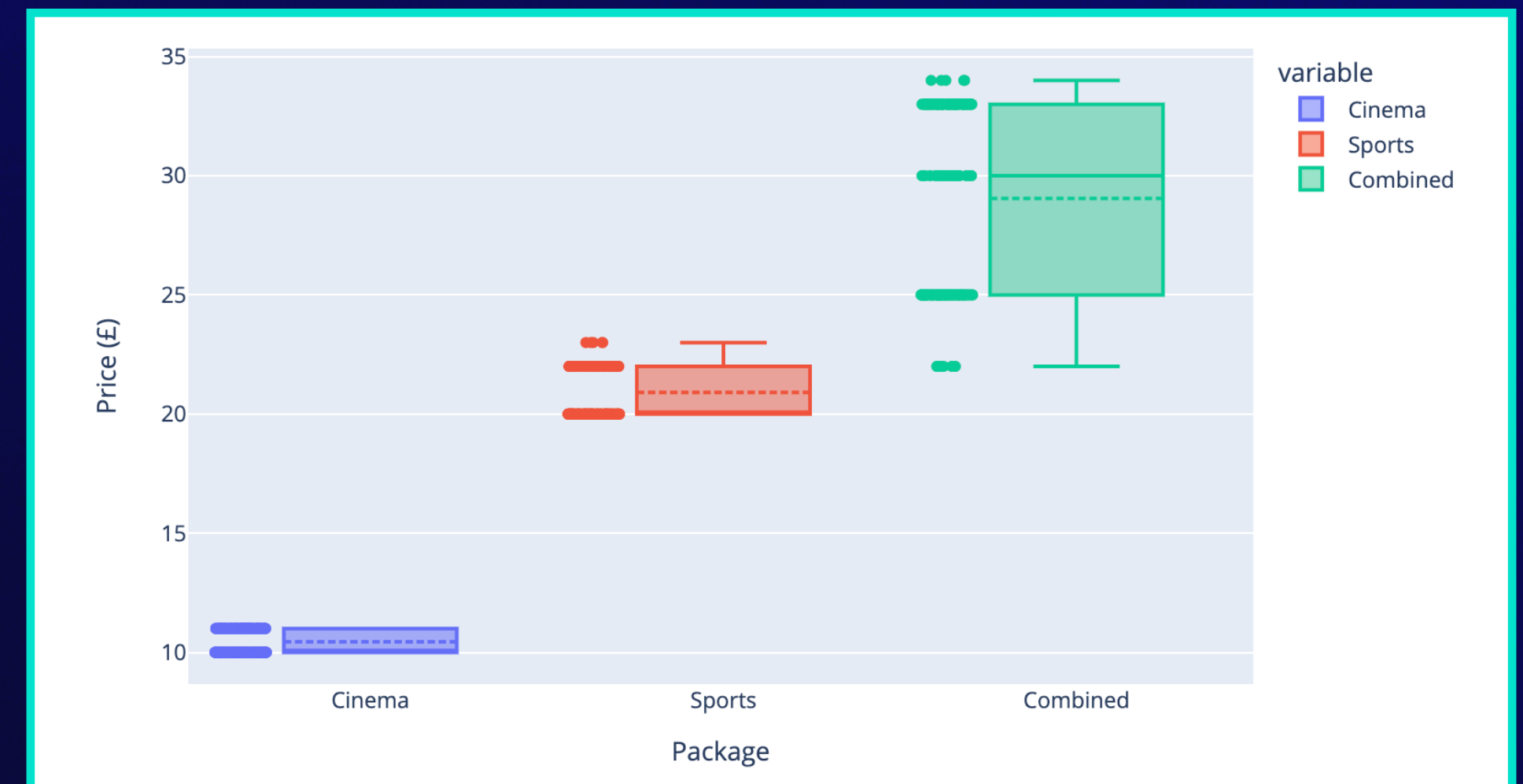
Summary of cleaning

- *The data was quite clean.*
 - no null values
 - no duplicates
 - eleven (11) unique regions.
- *Potential outliers in the ‘Sales’ data* are to be investigated further.

Box Plot for Sales Figures



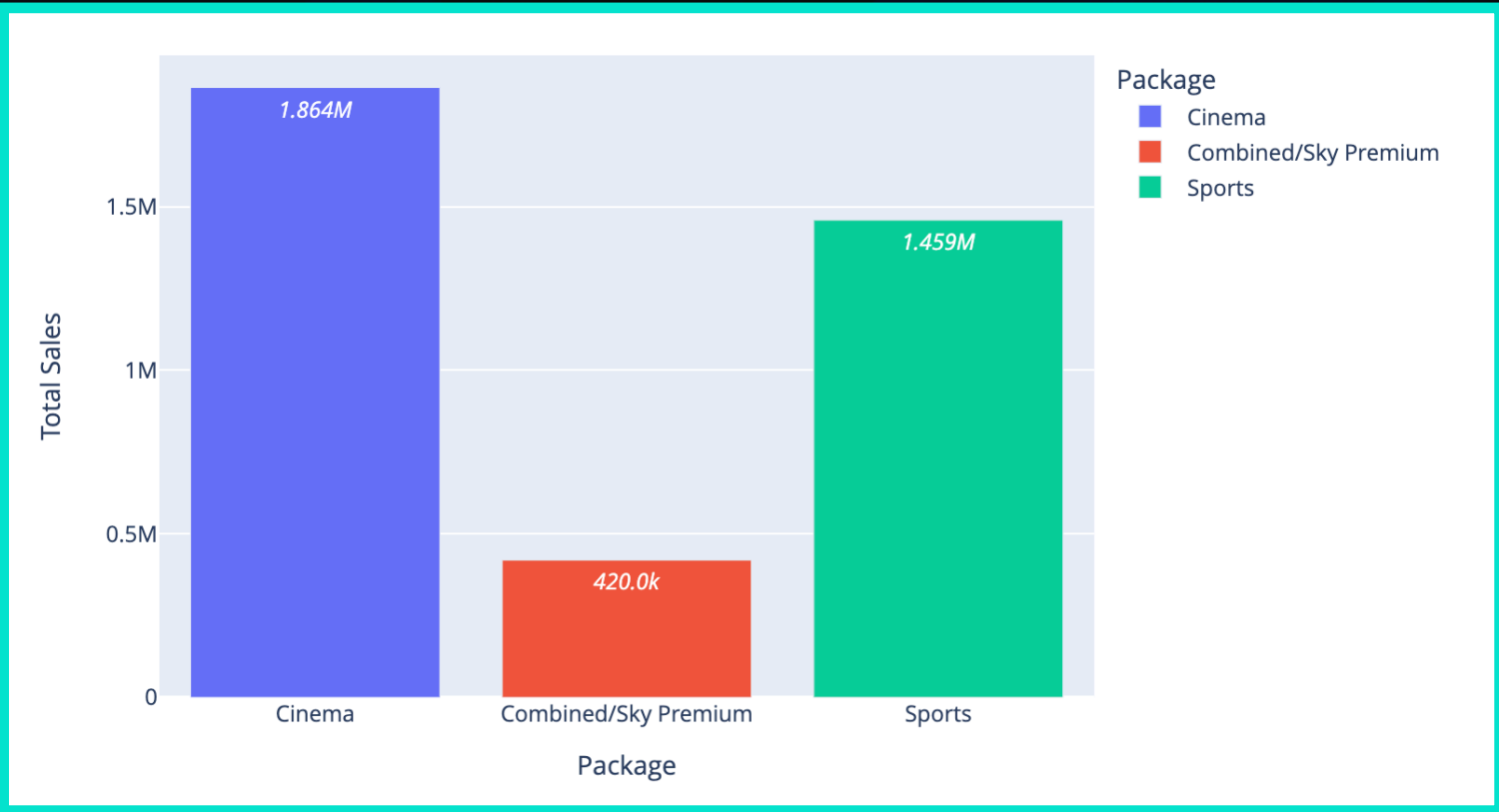
Box Plot for Cinema, Sports, and Combined Packages



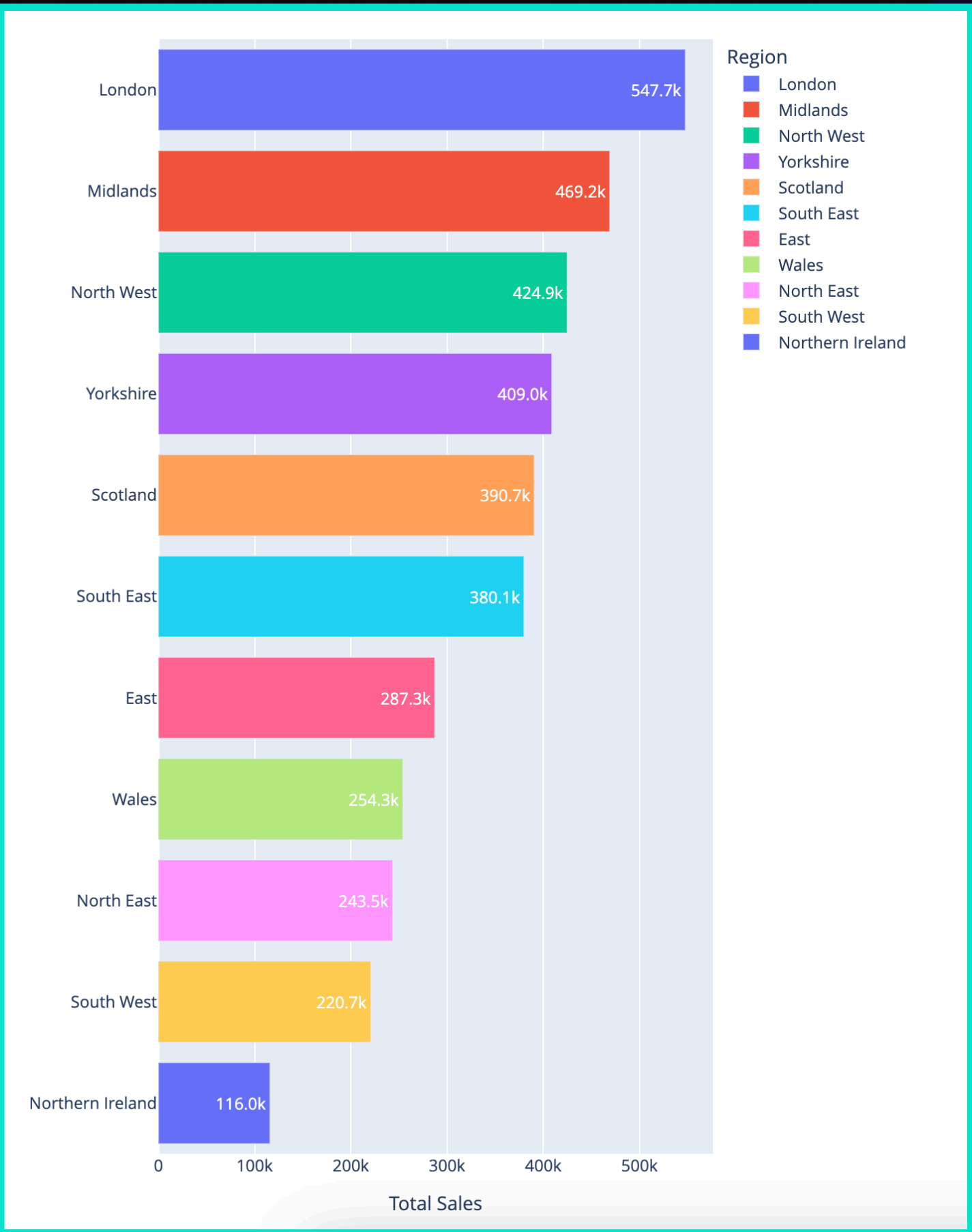
Exploratory Data Analysis (EDA).

- We check the **Total Sales/Revenues by Package** and **Region** and **Weekly Sales Volumes**.

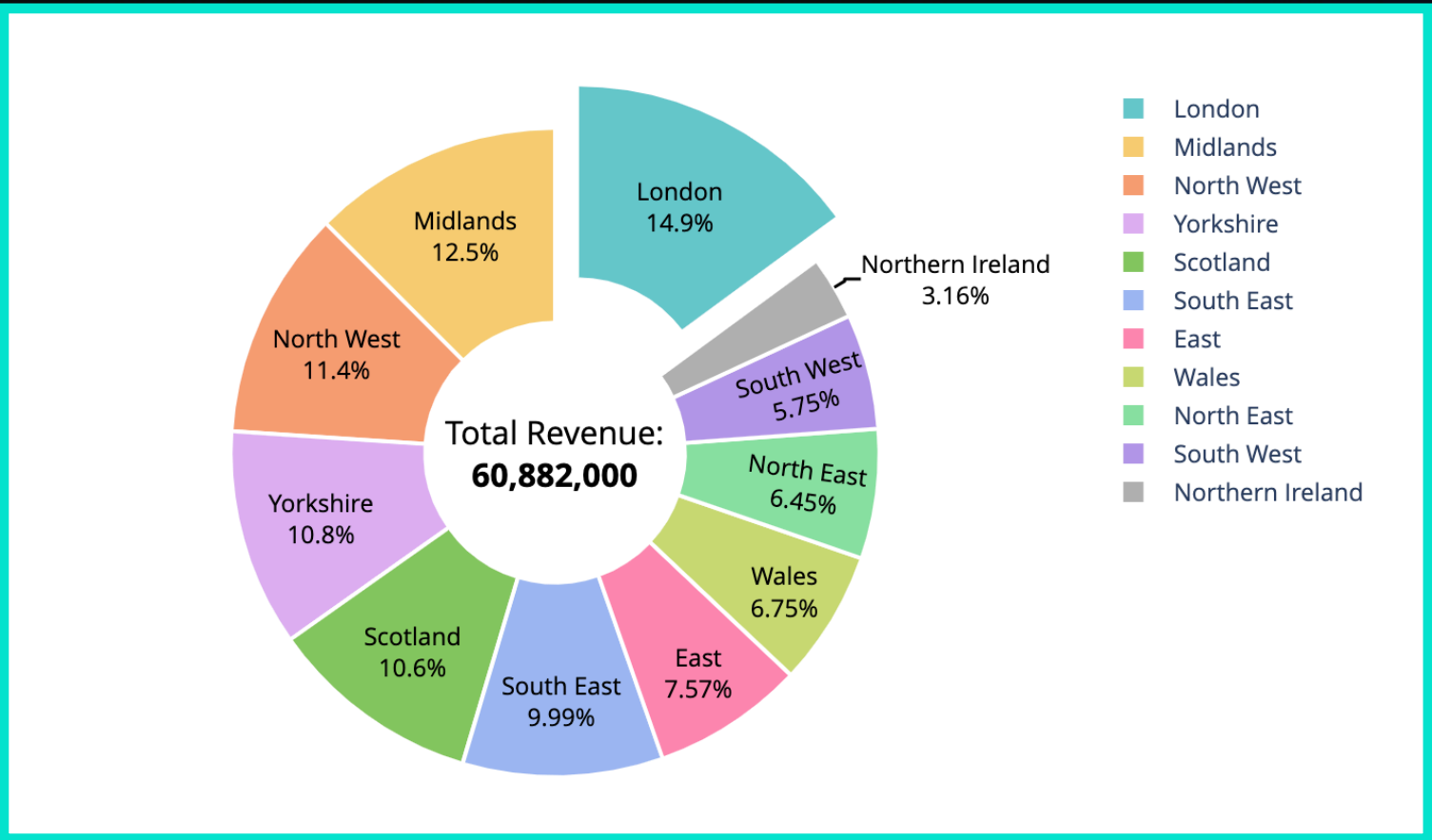
Total Sales by Package



Total Sales by Region



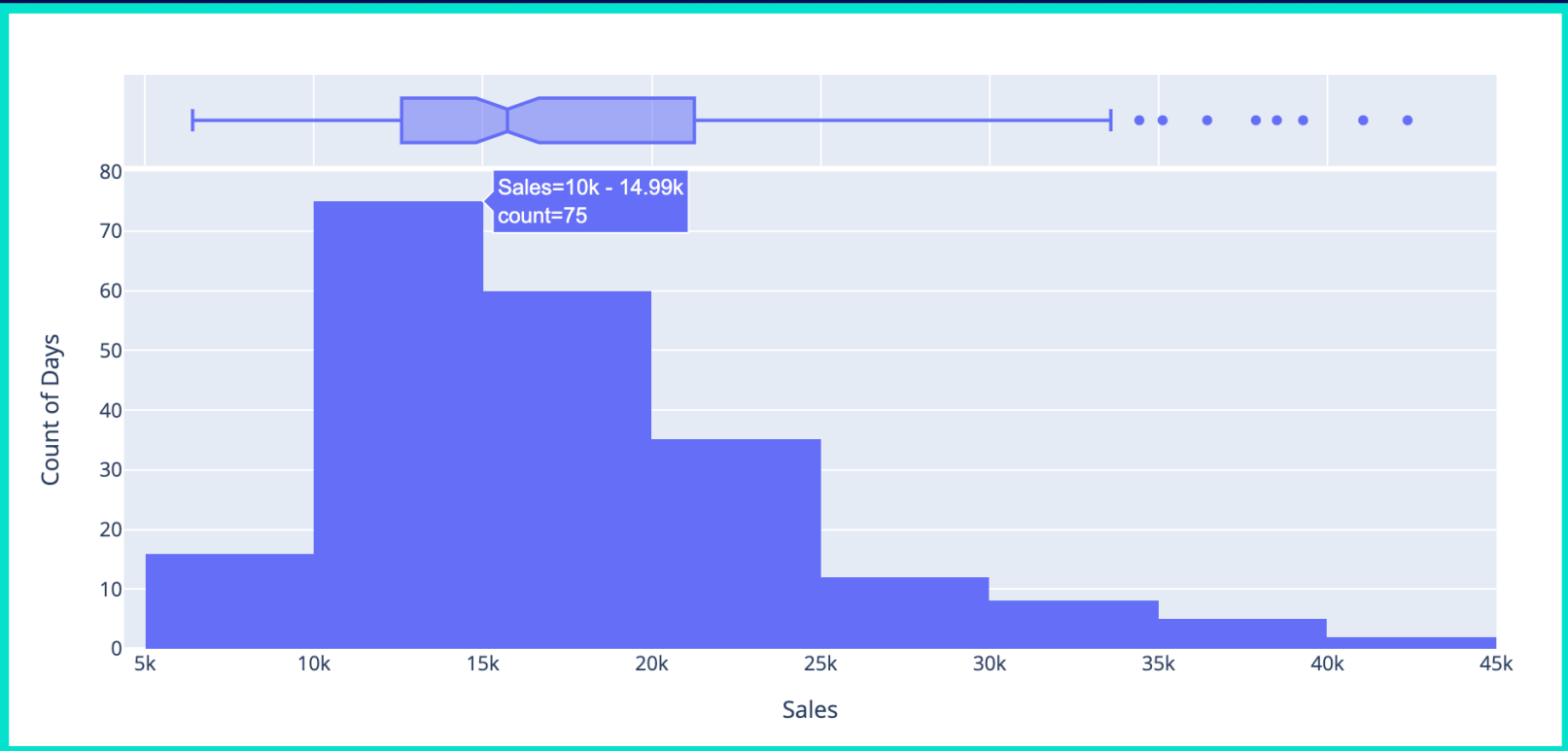
Total Revenue by Region



Total Revenue by Package



Distribution of Total Weekly Sales

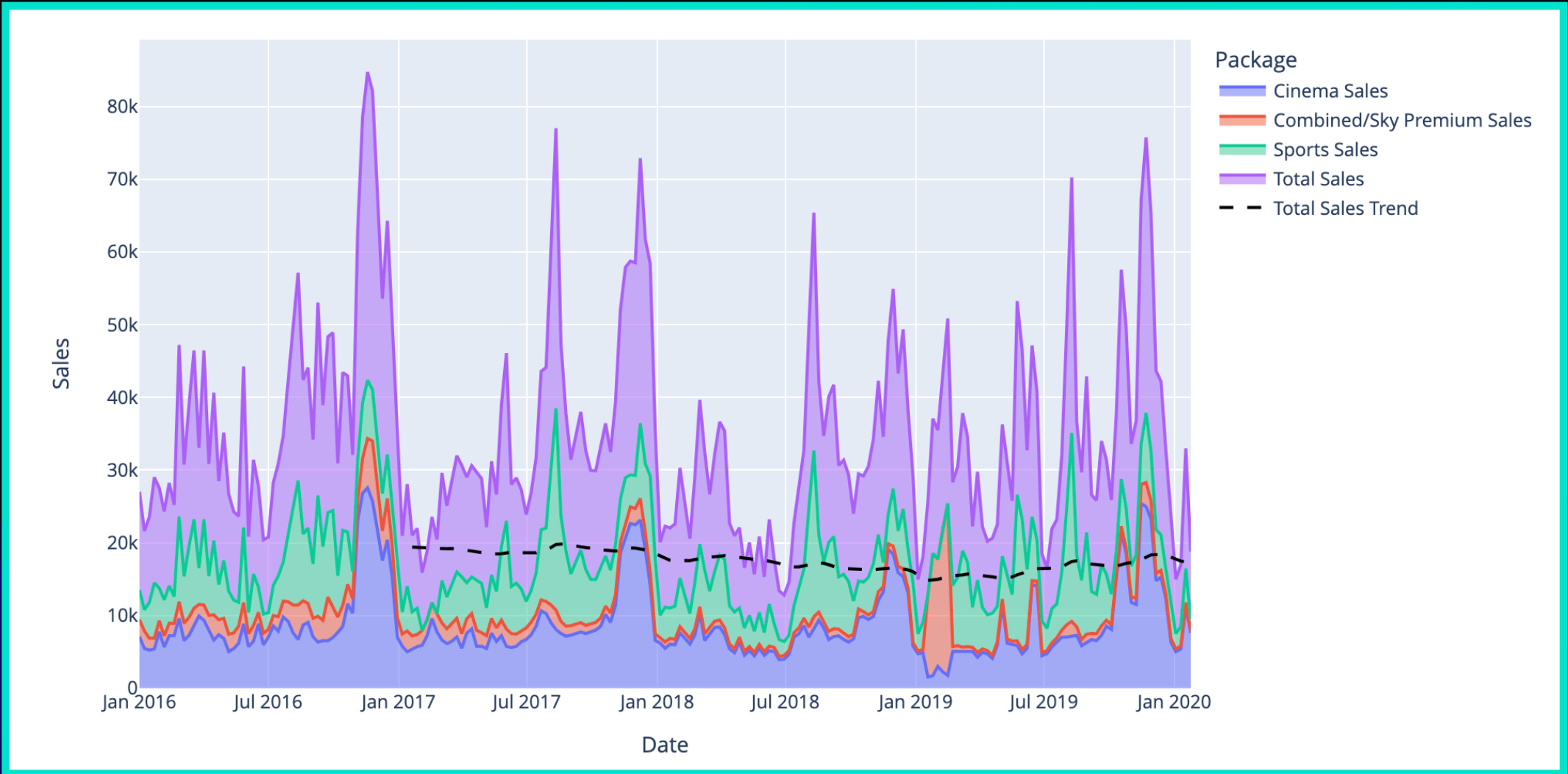


Package Sales Analysis.

- Let's examine the descriptive statistics for Weekly Sales grouped by **Package**.
- We see an oddity: **Nine (9) weeks were no “Sport” packages sold.** Premium and Cinema Packages.
- Weeks with missing Sports sales are:

3 February 2017	18 January 2019
10 February 2017	25 January 2019
17 February 2017	1 February 2019
24 February 2017	8 February 2019
	15 February 2019

Sales Evolution by Package



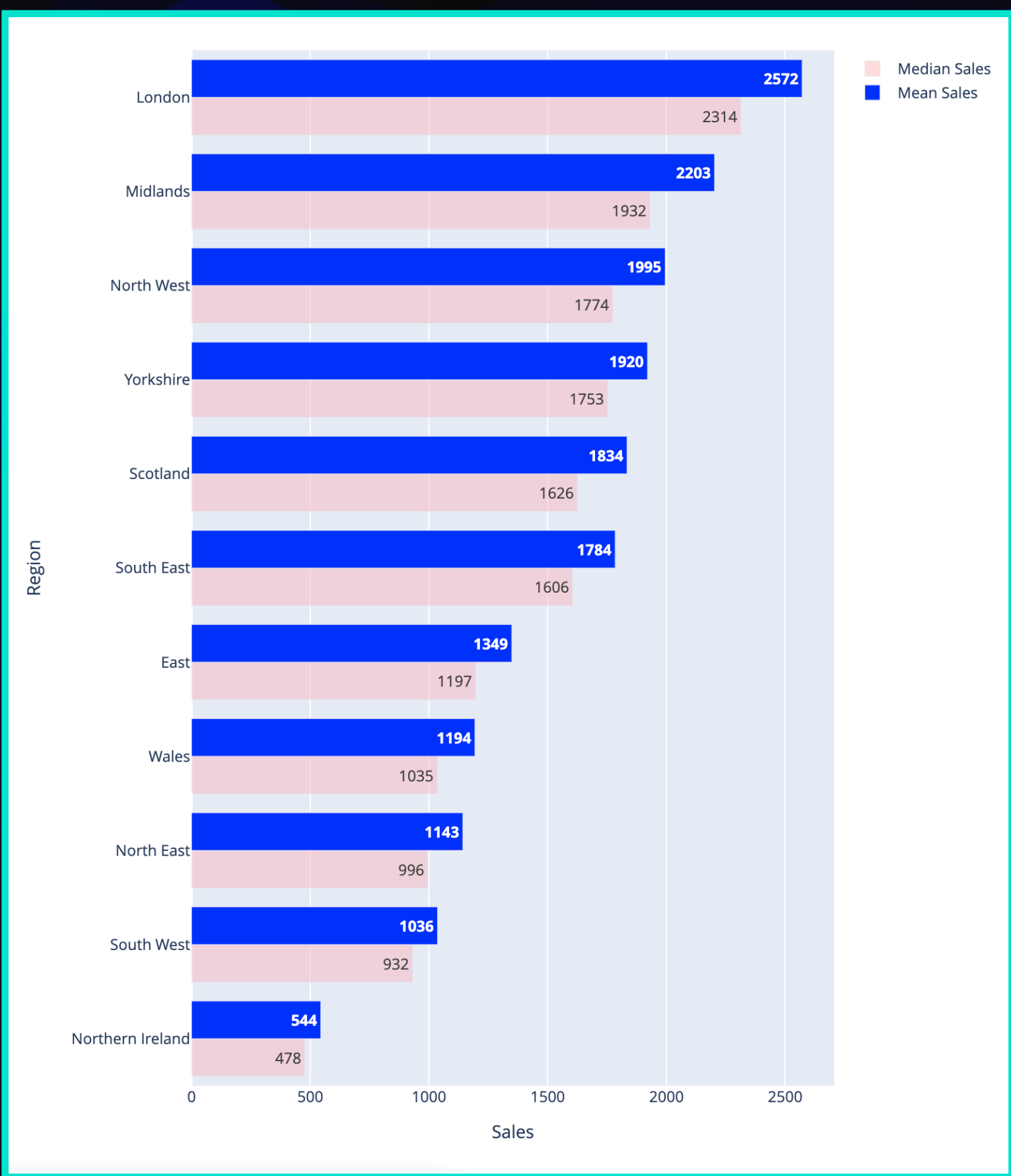
Descriptive Statistics of Weekly Sales by Package

	Sky Premium	Sports	Cinema	Total Sales
count	213.00	204.00	213.00	213.00
mean	1,971.80	7,152.57	8,752.49	17,574.64
std	2,735.08	4,104.64	5,035.13	7,132.78
min	263.00	1,498.00	1,543.00	6,411.00
25%	720.00	4,586.25	5,786.00	12,615.00
50%	1,380.00	5,995.00	7,160.00	15,727.00
75%	2,213.00	8,635.00	9,356.00	21,202.00
max	23,654.00	27,737.00	27,550.00	42,372.00

Regional Sales Analysis.

- Here, we examine the *mean* and *median* sales per *region*.
- We use measures of *variability* to see which regions have the most *stable/predictable* sales.
- Observe weeks with the *highest sales numbers* for regions recording weekly sales **above 4 thousand**.

Mean and Median Weekly Sales by Region



Region	Mean	CV (%)
Wales	1,194	53.73
Northern Ireland	544	48.93
North East	1,143	47.87
Scotland	1,834	46.51
North West	1,995	44.44
Midlands	2,203	43.93
South West	1,036	43.19
East	1,349	42.61
Yorkshire	1,920	41.98
London	2,572	41.02
South East	1,784	40.47

Region	Week with Highest Sales	Sales on that Date
London	Friday, May 24, 2019	7647
Wales	Friday, June 02, 2017	4975
Scotland	Friday, April 15, 2016	6193
North West	Friday, August 11, 2017	5656

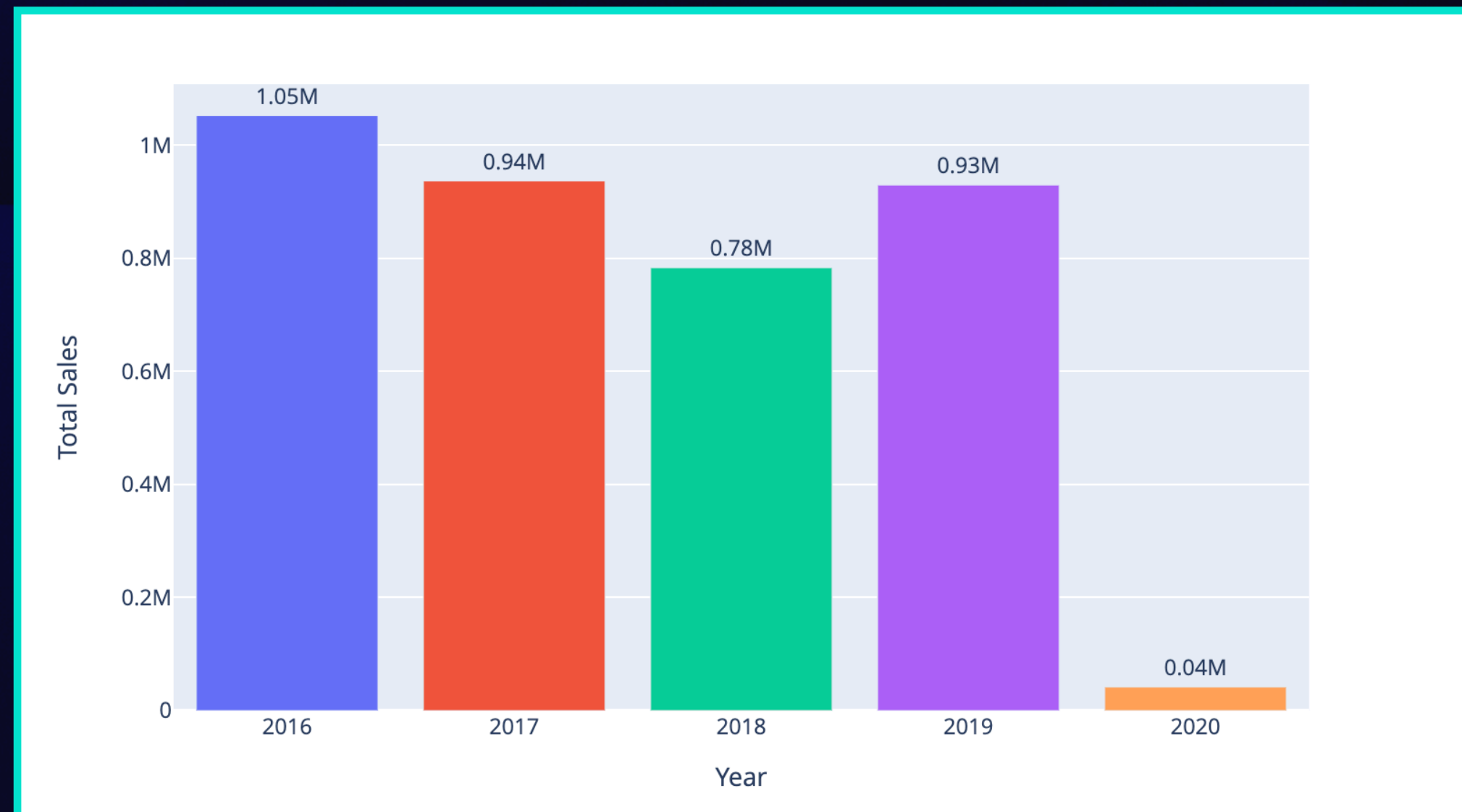
Weekly Sales Distribution by Region



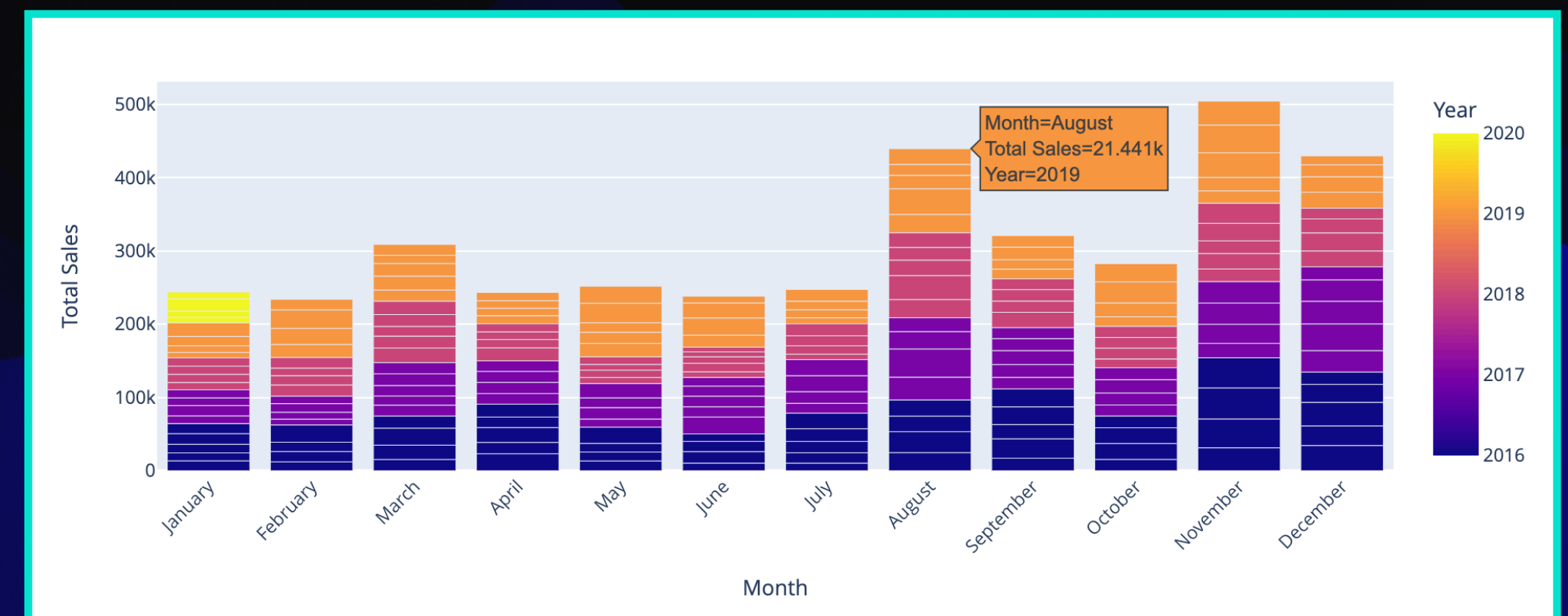
Temporal Analysis.

- Here, we take a look at the **total sales by year**, the **total sales by month** and the **top 10 highest-selling weeks**.

Total Sales by Year



Total Sales by Month



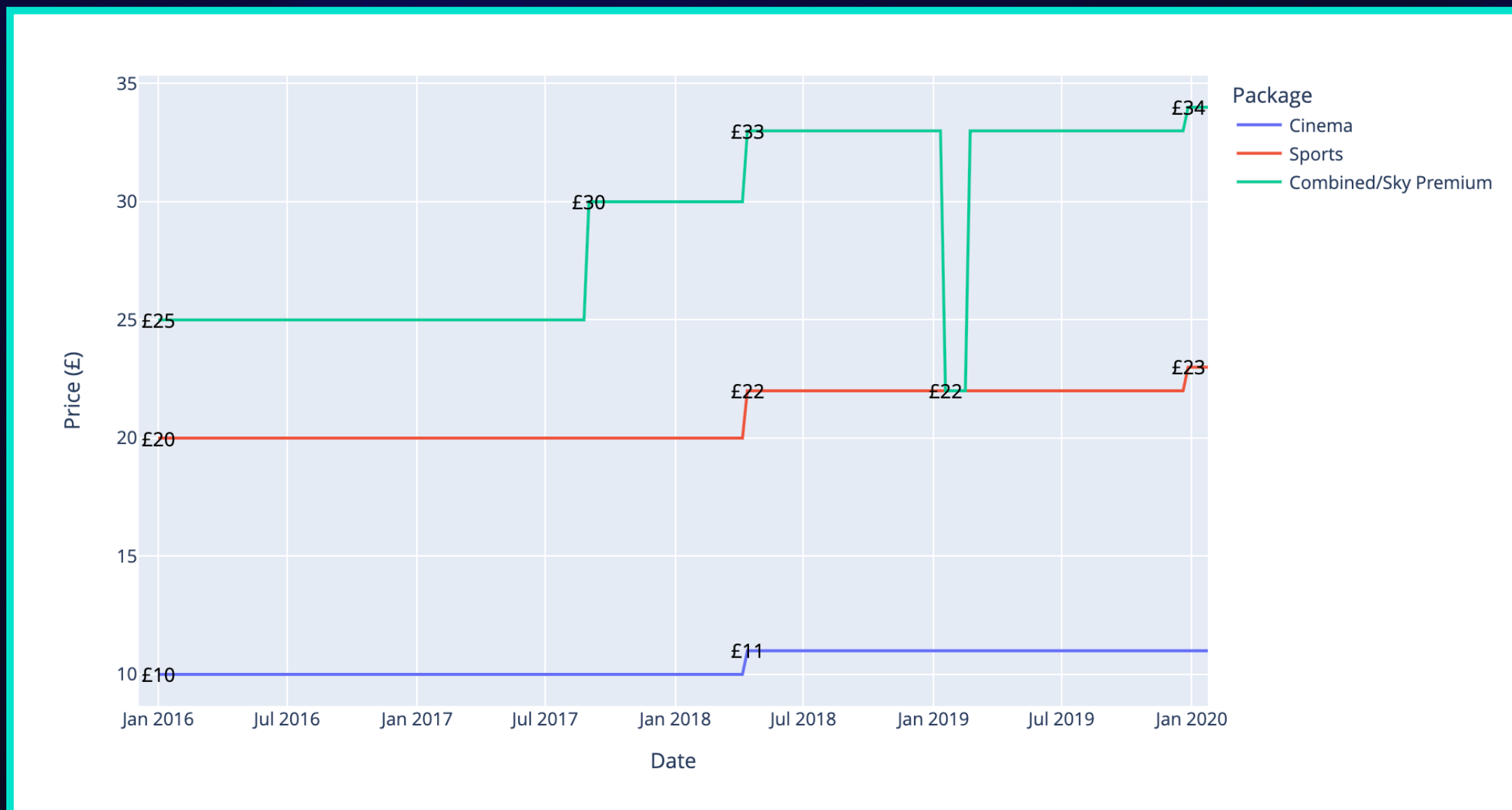
Top 10 Weeks with the Highest Sales



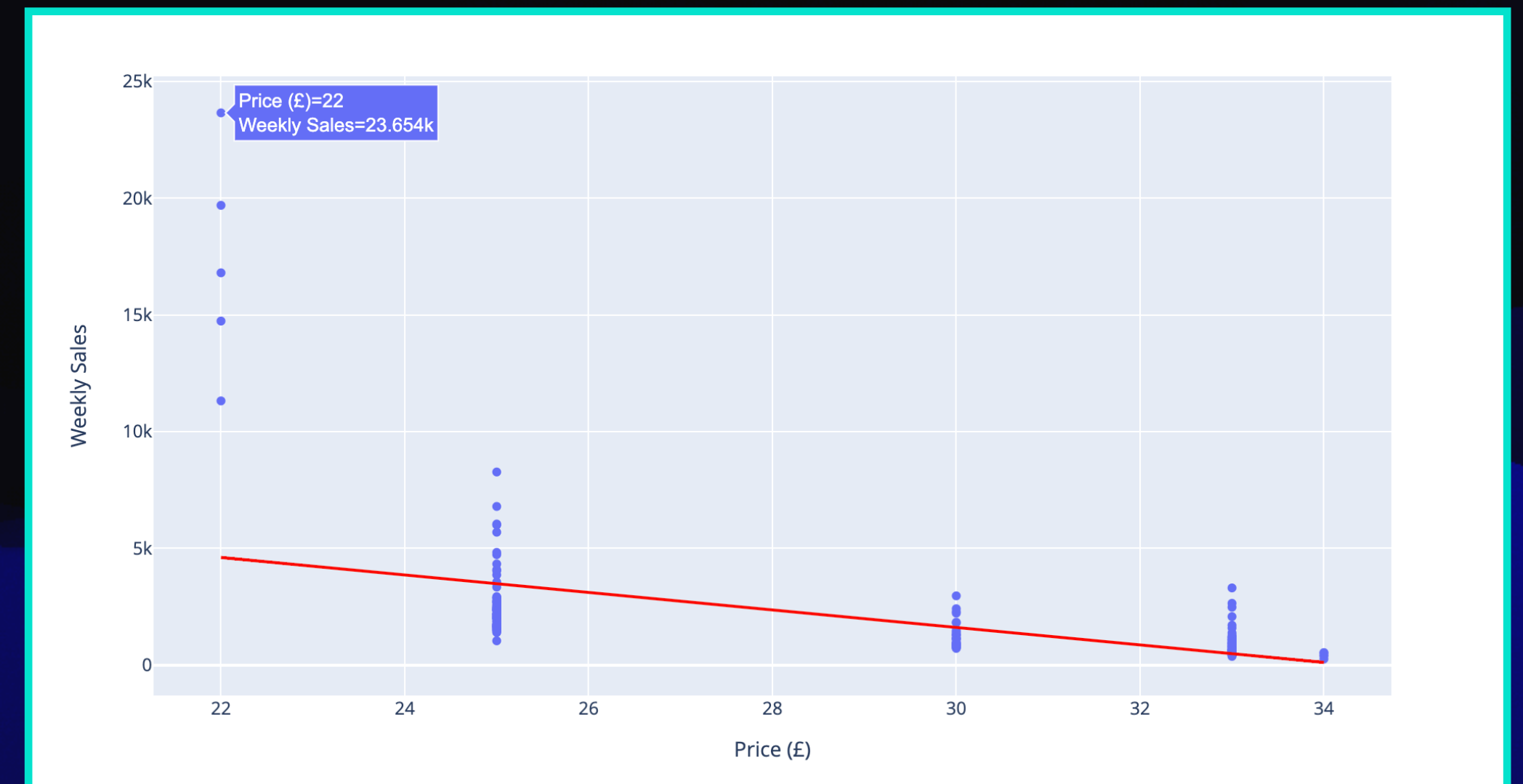
Price Analysis.

- Prices for Sports and Cinema packages are stable over time.
- Sky Premium's price saw the biggest changes.
- On **April 13 2018**, there was an across-the-board price increase for all packages.
- During **Jan & Feb 2019**, Sky Premium was priced the same as the Sports package.
- Now, we look at the relationship between price and daily sales for Premium.

Progression of Package Prices Over Time



Weekly Sales vs Prices for Combined Package



- Daily Sales and the Premium package price seem to be **negatively related**.
- The relationship appears to be **non-linear**.
- Next, we have a look at a **regression model** taking these hypotheses into account.

Premium Price Impact on Sales: Regression Analysis.

$$\text{Weekly_Premium_Sales} = \beta_0 + \beta_1 \cdot \text{Price} + \beta_2 \cdot \text{Price}^2 + \varepsilon$$

	Coefficient	Std. Error	t-value	p-value	Significance
Intercept	165,723.314	44,627.530	3.713	0.000	***
Price	-11,192.343	3,107.153	-3.602	0.000	***
Squared Price	187.956	53.241	3.530	0.000	***
R-squared	0.547				
Adj. R-squared	0.543				
F-statistic	47.679				***
No. Observations	213.000				
Note	[1] Standard Errors are heteroscedasticity robust (HC3)				
	*** = the coefficients are statistically significant at the 1% level.				

- Increasing the premium package price leads to a decrease in weekly sales, but the rate of decrease diminishes as the price rises further.
- This suggests a potential pricing sweet spot where revenue can be maximised.

Addressing the Data Anomaly: The missing Sports Sales.

Year Missing Weeks occurred	Average Weekly Revenue for the Year	Average Revenue of the Missing Weeks	Min % Change	Max % Change	Avg % Change
2017	281,968	131,949	-60.92%	-46.39%	-53.2%
2019	309,390	401,993	-14.04%	74.50%	29.93%

2017 Observations:

- All the given dates in 2017 have revenue values that are **below the average for the year**.
- The percentage changes range between around -46% to -61%.
- This suggests that the revenues on these particular **weeks were considerably lower** than what was **typically** observed for the **year**.
- If the **sports package was unavailable** during this period, it provides a plausible explanation for the reduced revenue.

2019 Observations:

- The given dates in 2019 show a **mixed trend**. Some dates have revenues below the yearly average, while others are above.
- The percentage changes range from -14% (a decrease) to about +74% (an increase).
- The dates with increased revenue (particularly the latter three) show **a significant jump**, suggesting some form of boost in sales.
- Here sports sales were missing where there was a simultaneous reduction in the combined package. **Indicating a Special Offer**.
- Customers' shift to purchasing the combined package **effectively boosted revenue**. It's impressive since February is a slower month.

Commercial Insights & Actionable Takeaways.

1. **Package Sales:** A longer re-run of the Premium offer and customer feedback surveys can give more insights into the robustness of seasonal pricing on revenue.
 - 1.1. Promotions may boost sales during slower months.
2. **Pricing Trials:** Trials/Simulations with varying package prices will help forecast sales/revenues more accurately.
3. **Unpack Sales Spikes:** It is worth looking into the external/internal events or factors that are leading to the sales/revenue spikes in certain months and weeks.
4. **Regional Variation:** The data suggests that some regions have more predictable sales patterns than others; it could also be beneficial to do a deep dive into the demand patterns of these regions.