Investigating Ticket Sales Data with **Amazon Redshift**

Overview

In this workspace, we will be accessing data stored in Amazon Redshift, a data warehouse product that is part of Amazon Web Services. More specifically, we'll be analyzing sales activity from a fictional ticketing website where users both buy and sell tickets online for sporting events, shows, and concerts **(source).**

Problem Statement

This script analyzes sales data from a fictional ticketing website to understand user behavior, sales trends, and identify potential areas for improvement.

Goals

- Explore event information, including location, category, and scheduling.
- Analyze listing and sales data to understand how many tickets are sold per listing and identify top sellers and buyers.
- Examine sale prices across different event categories.
- Find listings with inconsistencies, such as sales happening before the listing was created.
- Identify users who might benefit from advertising their listings based on unrealized sales potential.
- Visualize sales trends over time by category.

Dataset & Source

The data resides in an Amazon Redshift data warehouse, containing tables for:

- **Venue:** Details about event locations.
- **Category:** Event categories (sports, shows, concerts etc.).
- **Date:** Dates of events.
- **Event:** Information about events, including venue, category, and start time.
- **Listing:** Information about ticket listings, including listing time, seller ID, price, and number of tickets.
- **Sales:** Information about ticket sales, including buyer ID, quantity sold, and sale time.
- **Users:** User information like username, name, and city.

Data Findings

- Most events occur in the same city as the venue.
- The majority of listings have only one associated sale.
- A small number of users sold a significant number of tickets in 2008.
- There are variations in average and median ticket prices across event categories.
- Some listings have sale times before the listing was created (considered bad data).
- The script identifies users with a high potential for unrealized sales (listings that could have sold but didn't).
- Sales trends over time can be visualized by category.

The number of events hosted in different cities

Key Observations:

- **New York City dominates:** With 2647 events, it clearly stands out as the city hosting the most events.
- **Los Angeles and Las Vegas follow:** These two cities have the second and third highest number of events, indicating a strong presence of events in the entertainment and sports industry.
- **Chicago and San Francisco:** These cities also host a significant number of events, suggesting a vibrant cultural and entertainment scene.
- **Lower Tier Cities:** Cities like Montreal, Newark, Irving, Sunrise, and Calgary host a much smaller number of events compared to the top cities.

Overall:

The table highlights the concentration of events in major cities, particularly New York City, Los Angeles, and Las Vegas. This suggests that these cities have a strong demand for live events and are well-equipped to host a variety of events throughout the year.

	venuecity	num_event
0	New York City	2647
1	Los Angeles	312
2	Las Vegas	300
3	Chicago	209
4	San Francisco	194
...	...	...
74	Montreal	27
75	Newark	27
76	Irving	25
77	Sunrise	24
78	Calgary	18

The distribution of listings based on the number of sales they have generated

Here's a breakdown:

- 48,029 listings have only been sold once. This is the most common scenario.
- 36,570 listings have been sold twice.
- 14,665 listings have been sold three times.
- 1,808 listings have been sold four times.
- Only 12 listings have been sold five times.

	number_of_sales	cnt
0	1	48029
1	2	36570
2	3	14665
3	4	1808
4	5	12

From this data, we can infer that:

- Most listings on the platform have a relatively low number of sales.
- A small number of listings generate a higher number of sales.
- The distribution is skewed towards listings with fewer sales, indicating that it's common for listings to have limited success.

This information can be useful in understanding the overall sales patterns on the platform and identifying areas for potential improvement, such as optimizing listing visibility or implementing strategies to encourage repeat sales.

The top 5 sellers based on the number of tickets sold (represented by the "sum" column)

	sellerid	username	name	city	sum
0	48950	TUT90BHI	Nayda Hood	Frisco	46
1	20029	RPM45HGY	Drew Mcguire	Lancaster	41
2	19123	DZW00VOQ	Scott Simmons	Carson	41
3	36791	DCE77DOA	Emerson Delacruz	Springfield	40
4	41579	QXQ28WLA	Harrison Durham	Anaconda	39

Here's a breakdown:

- Nayda Hood (TUT90BHI) is the top seller with 46 tickets sold.
- Drew Mcguire (RPM45HGY) and Scott Simmons (DZW0OVOQ) are tied for the second spot with 41 tickets sold each.
- Emerson Delacruz (DCE77DOA) sold 40 tickets.
- Harrison Durham (QXQ28WLA) sold 39 tickets.

This data can be used to identify top-performing sellers and potentially reward or incentivize them. Additionally, it can help understand user behavior and identify potential areas for improvement in the platform's features or marketing strategies.

-

The table shows the top 5 buyers based on the number of tickets purchased (represented by the "sum" column)

	buyerid	username	name	city	sum
0	8933	CNF70VPH	Jerry Nichols	Middlebury	67
1	1298	EDB46JXK	Kameko Bowman	Newburyport	64
2	3797	KTV94TWB	Armando Lopez	Pomona	64
3	5002	CBC51API	Kellie Savage	Falls Church	63
4	4064	IQS59DPH	Kadeem Blair	Anaheim	60

Here's a breakdown:

- Jerry Nichols (CNF70VPH) is the top buyer with 67 tickets purchased.
- Kameko Bowman (EDB46JXK) and Armando Lopez (KTV94TWB) are tied for the second spot with 64 tickets purchased each.
- Kellie Savage (CBC51API) purchased 63 tickets.
- Kadeem Blair (IQS59DPH) purchased 60 tickets.

This data can be used to identify top-performing buyers and potentially reward or incentivize them. Additionally, it can help understand user behavior and identify potential areas for improvement in the platform's features or marketing strategies.

The average and median ticket prices for two event categories: Concerts and Shows.

Key Observations:

- **Average Ticket Price:**
  - Concerts have an average ticket price of $333.76.
  - Shows have an average ticket price of $336.98.
- **Median Ticket Price:**
  - Concerts have a median ticket price of $229.
  - Shows have a median ticket price of $232.

	catgroup	avg_ticket_price	median_ticket_price
0	Concerts	333.755006	229.0
1	Shows	336.982704	232.0

Insights:

- **Similar Average Prices:** The average ticket prices for Concerts and Shows are very close, suggesting that, on average, tickets for these events are similarly priced.
- **Median Price Differences:** The median ticket price for Shows is slightly higher than for Concerts. This suggests that the middle price point for Shows is higher compared to Concerts.
- **Potential Price Disparity:** The difference between the average and median prices for both categories suggests that there might be a significant number of higher-priced tickets in both categories, potentially skewing the average price upwards.

Are there listings where the sale happened before the listing?

	count
0	2965

The table shows that there are 2965 listings where the sale happened before the listing. This is likely due to errors in the data or potential system issues.
It's important to investigate and clean this data before further analysis, as it could lead to misleading results and incorrect conclusions.

Possible reasons for this inconsistency:

- **Data Entry Errors:** Mistakes during data input or processing could lead to incorrect timestamps.
- **System Time Zone Issues:** Differences in time zones between systems could cause discrepancies in timestamps.
- **System Bugs or Glitches:** Technical issues might have caused the system to record incorrect timestamps.

To address this issue, the following steps can be taken:

1. **Data Cleaning:** Remove or correct the listings with incorrect timestamps.
2. **Data Validation:** Implement stricter data validation checks to prevent such errors in the future.
3. **System Review:** Investigate the system for any potential bugs or issues that might be causing the problem.

The table shows the top 100 users with the highest unrealized sales. This means that these users had listings with a high potential for sales that were not realized.

	sellerid	name	unrealized_sales
0	25428	Jaime Wagner	58395.0
1	24896	Macey Ortiz	53086.0
2	49322	Dustin Vincent	50914.0
3	36926	Audrey Barber	50345.0
4	45819	Kelly Barrett	49826.0
...	...	...	...
95	35926	Ulysses Kinney	38895.0
96	45372	Lysandra Sanchez	38862.0
97	48188	Caesar Parrish	38847.0
98	25373	Jakeem Byrd	38847.0
99	1759	Lane Franklin	38740.0

What does this mean for the ticketing website?
By targeting these users with advertising suggestions, the website can help them sell their remaining tickets and generate more revenue. This is beneficial for both the users and the platform.

Here's how the website could leverage this information:

- **Targeted Advertising:** The website can offer targeted advertising options to these users, such as sponsored listings or email campaigns.
- **Personalized Recommendations:** The platform can provide personalized recommendations on how to improve their listings, such as using better titles, descriptions, or images.
- **Incentives:** The website could offer incentives to these users, such as discounts on listing fees or commission reductions, to encourage them to advertise their listings.

By taking these steps, the website can help these users maximize their sales potential and ultimately drive more revenue for the platform.

The total sales for Concerts and Shows for each week of the year 2008

Key Observations:

- **Sales Fluctuate:** The total sales for both Concerts and Shows fluctuate from week to week. There are weeks with high sales and weeks with lower sales.
- **Seasonal Trends:** While the data doesn't cover a full year, we can see potential seasonal trends. For example, there might be higher sales during certain seasons or around specific holidays.
- **Category Comparison:** We can compare the sales of Concerts and Shows week by week to see which category is performing better.

Potential Insights:

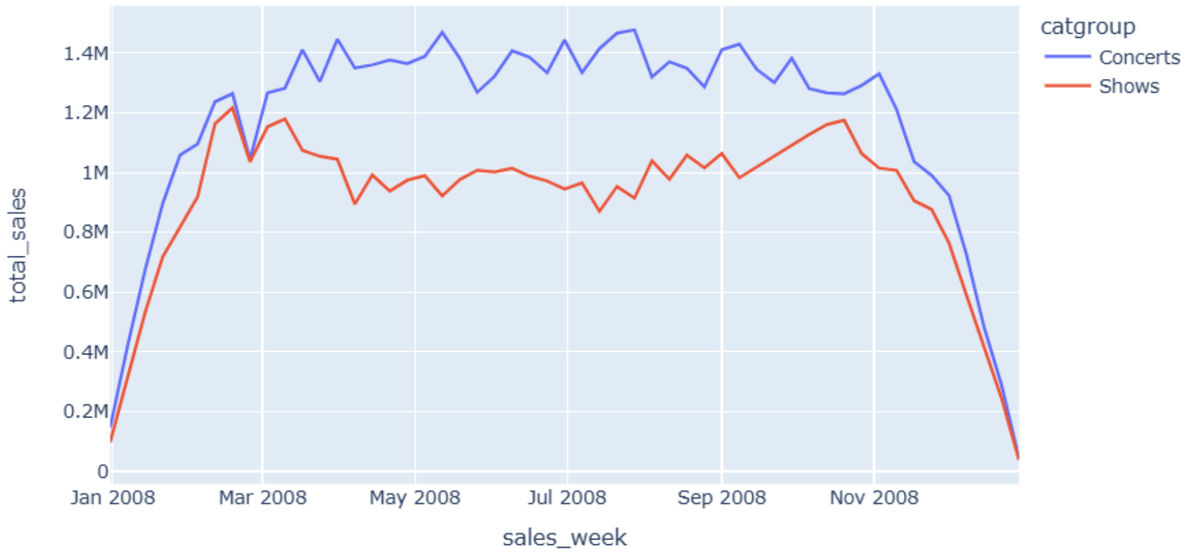
- **Identify Peak Sales Periods:** By analyzing the data, we can identify the weeks with the highest sales for each category. This information can be used to optimize marketing efforts and inventory management during these peak periods.
- **Understand Seasonal Trends:** If there are clear seasonal patterns, businesses can adjust their pricing strategies, marketing campaigns, and inventory levels accordingly.
- **Compare Category Performance:** By comparing the sales of Concerts and Shows, businesses can gain insights into consumer preferences and adjust their offerings accordingly. For example, if Shows are consistently outperforming Concerts, they may want to invest more in promoting and developing Show-related events.

	catgroup	sales_week	total_sales
0	Concerts	2007-12-31 00:00:00+00:00	146716.0
1	Shows	2007-12-31 00:00:00+00:00	98275.0
2	Concerts	2008-01-07 00:00:00+00:00	420710.0
3	Shows	2008-01-07 00:00:00+00:00	313664.0
4	Shows	2008-01-14 00:00:00+00:00	532037.0
...	...	...	...
101	Shows	2008-12-15 00:00:00+00:00	419135.0
102	Shows	2008-12-22 00:00:00+00:00	245468.0
103	Concerts	2008-12-22 00:00:00+00:00	289723.0
104	Shows	2008-12-29 00:00:00+00:00	37838.0
105	Concerts	2008-12-29 00:00:00+00:00	44918.0

The total sales for Concerts and Shows for each week of the year 2008

Key Observations:

- **Seasonal Fluctuations:** We can observe distinct seasonal patterns in the sales data. Both Concerts and Shows exhibit peaks and troughs throughout the year.
- **Concert Dominance:** Overall, Concerts seem to have higher sales compared to Shows. This is evident in the higher peaks and overall trend of the blue line (Concerts).
- **Similar Patterns:** Both categories follow a similar trend, with peaks and troughs occurring at roughly the same time. This suggests that external factors like holidays, seasons, or economic conditions might be influencing both categories.



Possible Explanations:

- **Holiday Seasons:** The peaks in sales might coincide with holiday seasons or school breaks, when people have more time and disposable income for entertainment.
- **Weather:** Weather conditions can impact outdoor events like concerts, leading to fluctuations in sales.
- **Economic Factors:** Economic factors like unemployment rates and consumer confidence can influence spending on entertainment.

Insights

- **Understanding user behavior (selling/buying patterns) and sales trends** can inform marketing strategies.
- **Analyzing ticket prices by category** helps identify potentially underpriced listings.