# ADVENTUREWORKS SALES ANALYSIS

(for the sales management team)

By Xuehong Liu

The goal of this report is to assist the Adventureworks sales management team to develop data-driven management strategies moving forward. Sales are analyzed by territory, with drill-downs into orders by store and the top-selling sales representative at each store.

---

## 1. Summary of sales from last year and Sales YTD by territory

Fig. 1 is a comparison of total sales from last year and sales year-to-date (YTD) by territory. The territories of interests include Northwest, Northeast, Central, Southwest, Southeast, Canada, France, Germany, Australia and United Kingdom.

It can be seen that the highest sales occurred in Canada and Southwest last year (Fig. 1 left panel), each accounting for over 15% of the total sales. Sales YTD in Southwest outpaced all of the other territories, accounting for nearly 20% of the total sales (Fig. 1 middle panel). United Kingdom, Germany and Australia had the most increase in Sales YTD from last year (Fig. 1 right panel) where they already surpassed sales from last year by over 2.5 times. On the other hand, Sales YTD from Northeast and Southeast are only little over half of the total sales from last year.

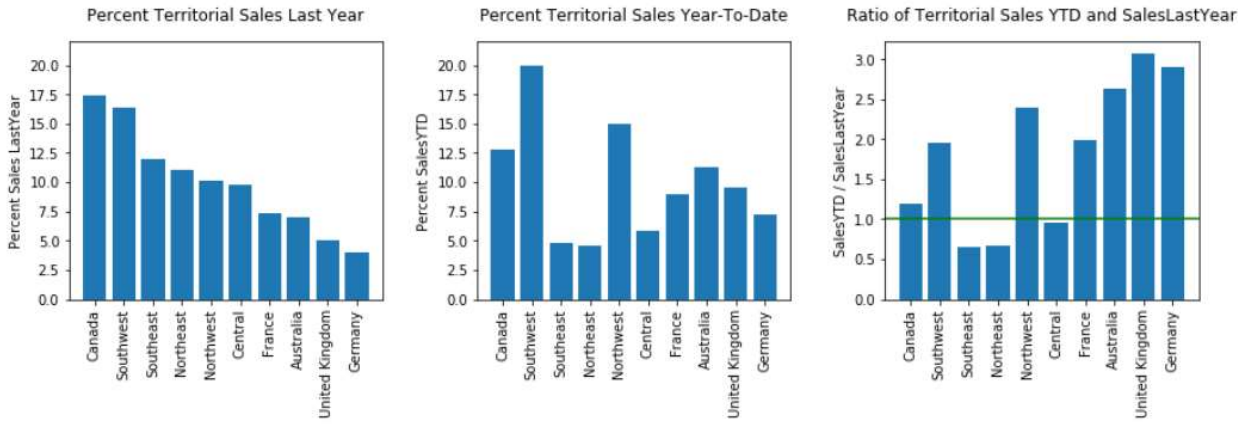


**Fig. 1. Summary of sales from last year and sales year-to-date (SalesYTD) by territories.** Left: Percent total sales from last year in each territory in respect to total sales last year. Middle: Percent SalesYTD in each territory in respect to total SalesYTD. Right: Ratio of sales from last year and SalesYTD by territories. A ratio of one is represented by the green horizontal line.

## 2. Discovery of a major irregularity in the data set.

In order to learn about how the sales representatives performed in reseller sales, sales from each store were retrieved and summarized based on sales by each sales representative. The sales representatives included in this analysis are: Michael, Linda, Jillian, Garrett, Tsvi, Pamela, Shu, José, David, Lynn, Rachel, Jae and Ranjit.

Careful examination of the results revealed that total sales at different stores by the same sales representative are all the same, an unlikely scenario in the real world. This data is depicted in

Fig.2 using a heatmap. It can be seen that sales by each sales representative are all the same for all the stores where they made sales, as indicated by the same level of intensity for the colored bars for all the stores.

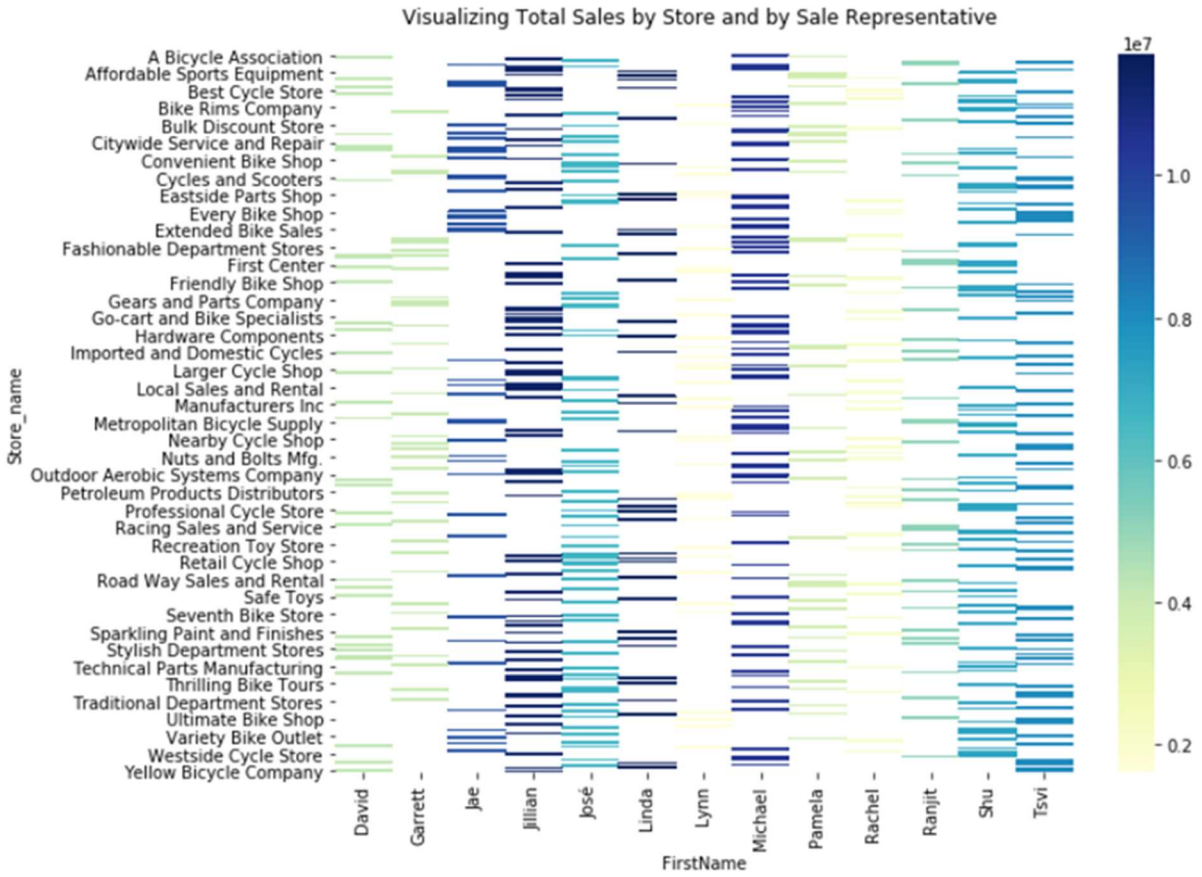


**Fig. 2. Summary of total store sales by sales representatives using a heatmap.** Store names are on the y-axis, sales representatives' names are on the x-axis. The sales by the same sales representative at all stores are coded with the same color. The total sales in each store are represented by the intensity of the colored bars. Darker the bar, higher is the sale.

## 3. Investigating the integrity of the data set.

Given the abnormality seen in the data set. Efforts were made to validate the data. As summarized in Table 1, the summary statistics for store sales by each sales representative suggest that the total sales in all the store by each sales representative are all the same. For example, David had sales in 38 stores where the minimum sales value, the maximum sales value and the other metrics are all the same. The near zero values for the standard deviation also suggest a uniformity of the data among different stores. My hypothesis is that during data entry, someone mistakenly entered the same values for all the stores.

**Table 1. Summary Statistics for Store Sales by Each Sales Representative**

| FirstName | count | mean | std | min | 25% | 50% | 75% | max |
|---|---|---|---|---|---|---|---|---|
| David | 38.0 | 4.207895e+06 | 9.438241e-10 | 4.207895e+06 | 4.207895e+06 | 4.207895e+06 | 4.207895e+06 | 4.207895e+06 |
| Garrett | 40.0 | 4.069422e+06 | 0.000000e+00 | 4.069422e+06 | 4.069422e+06 | 4.069422e+06 | 4.069422e+06 | 4.069422e+06 |
| Jae | 40.0 | 9.585125e+06 | 5.659122e-09 | 9.585125e+06 | 9.585125e+06 | 9.585125e+06 | 9.585125e+06 | 9.585125e+06 |
| Jillian | 76.0 | 1.134239e+07 | 1.125013e-08 | 1.134239e+07 | 1.134239e+07 | 1.134239e+07 | 1.134239e+07 | 1.134239e+07 |
| José | 74.0 | 6.683537e+06 | 5.626079e-09 | 6.683537e+06 | 6.683537e+06 | 6.683537e+06 | 6.683537e+06 | 6.683537e+06 |
| Linda | 39.0 | 1.169502e+07 | 9.434972e-09 | 1.169502e+07 | 1.169502e+07 | 1.169502e+07 | 1.169502e+07 | 1.169502e+07 |
| Lynn | 40.0 | 1.606441e+06 | 9.431870e-10 | 1.606441e+06 | 1.606441e+06 | 1.606441e+06 | 1.606441e+06 | 1.606441e+06 |
| Michael | 77.0 | 1.047537e+07 | 1.499887e-08 | 1.047537e+07 | 1.047537e+07 | 1.047537e+07 | 1.047537e+07 | 1.047537e+07 |
| Pamela | 38.0 | 3.748246e+06 | 2.359560e-09 | 3.748246e+06 | 3.748246e+06 | 3.748246e+06 | 3.748246e+06 | 3.748246e+06 |
| Rachel | 40.0 | 2.062393e+06 | 2.357968e-09 | 2.062393e+06 | 2.062393e+06 | 2.062393e+06 | 2.062393e+06 | 2.062393e+06 |
| Ranjit | 40.0 | 5.087977e+06 | 2.829561e-09 | 5.087977e+06 | 5.087977e+06 | 5.087977e+06 | 5.087977e+06 | 5.087977e+06 |
| Shu | 79.0 | 7.259568e+06 | 1.874547e-09 | 7.259568e+06 | 7.259568e+06 | 7.259568e+06 | 7.259568e+06 | 7.259568e+06 |
| Tsvi | 80.0 | 8.086074e+06 | 5.623191e-09 | 8.086074e+06 | 8.086074e+06 | 8.086074e+06 | 8.086074e+06 | 8.086074e+06 |

* The number of stores that each sales representative had sales at is listed in the first column 'count'. The mean store sales are listed in the 'mean' column. The standard deviation is listed in the 'std' column. '25%', '50%' and '75%' columns list sales values that are 25%, 50% and 75% of the maximum sales value.

To demonstrate this data abnormality in detail, individual transactions in each store was randomly checked for sales representatives. As shown in Table 2, the individual transactions of sales made by sales representative Michael at three different stores were compared. The top ten transactions are shown. It can be seen that individual transactions for sales made by Michael in three different stores are identical. Similar analysis was also conducted for other sales representatives. This finding highlights the need for rechecking the original data in the database since it is extremely unlikely for a sales representative to make identical sales in so many different stores.

**Table 2. Top Ten Sales in Three Different Stores by Michael**

| Store_name | FirstName | TotalDue | Store_name | FirstName | TotalDue | Store_name | FirstName | TotalDue |
|---|---|---|---|---|---|---|---|---|
| Active Cycling | Michael | 165028.7482 | Grand Discount Store | Michael | 165028.7482 | Retail Sales and Service | Michael | 165028.7482 |
| Active Cycling | Michael | 120512.2708 | Grand Discount Store | Michael | 120512.2708 | Retail Sales and Service | Michael | 120512.2708 |
| Active Cycling | Michael | 118581.3375 | Grand Discount Store | Michael | 118581.3375 | Retail Sales and Service | Michael | 118581.3375 |
| Active Cycling | Michael | 116463.9102 | Grand Discount Store | Michael | 116463.9102 | Retail Sales and Service | Michael | 116463.9102 |
| Active Cycling | Michael | 113231.0188 | Grand Discount Store | Michael | 113231.0188 | Retail Sales and Service | Michael | 113231.0188 |
| Active Cycling | Michael | 110266.9588 | Grand Discount Store | Michael | 110266.9588 | Retail Sales and Service | Michael | 110266.9588 |
| Active Cycling | Michael | 106618.2647 | Grand Discount Store | Michael | 106618.2647 | Retail Sales and Service | Michael | 106618.2647 |
| Active Cycling | Michael | 102017.9004 | Grand Discount Store | Michael | 102017.9004 | Retail Sales and Service | Michael | 102017.9004 |
| Active Cycling | Michael | 100902.4370 | Grand Discount Store | Michael | 100902.4370 | Retail Sales and Service | Michael | 100902.4370 |
| Active Cycling | Michael | 99952.4380 | Grand Discount Store | Michael | 99952.4380 | Retail Sales and Service | Michael | 99952.4380 |

* The three stores, 'Active Cycling', 'Grand Discount Store' and 'Retail Sales and Service' were selected randomly.

## 4. Top performers in each store

It is an important metrics to evaluate top sales in each store by sales representatives. Unfortunately, among the total 701 stores in the data set, only two stores are linked to more than one sales representative, namely, Friendly Bike Shop and Sports Products Store. Linda made more sales at Sports Products Store than Ranjit did while Michael made more sales at Friendly Bike Shop than David did. Clearly, this result is also contaminated by the abnormality in the data discussed above. The data must be corrected before meaningful analysis can be carried out to find top performers in each store.

**Table 2. Store Sales involve More than two sales Representatives**

| Store_name | FirstName | TotalDue |
|---|---|---|
| Sports Products Store | Linda | 1.169502e+07 |
| Sports Products Store | Ranjit | 5.087977e+06 |
| Friendly Bike Shop | Michael | 1.047537e+07 |
| Friendly Bike Shop | David | 4.207895e+06 |

* Only two stores from a total of 701 stores has more than one sales representative made sales.