**Report**

**Introduction**

This report will visualise data through the utilisation of SQL and Python, in order to allow the board to gain more clarity around their data, by translating the dataset provided into a more visual context.

The aim of this report is to show the data in a digestible format that will help you draw conclusions about your company and help drive business decisions backed by evidence.

The report will enlighten you on the following questions:

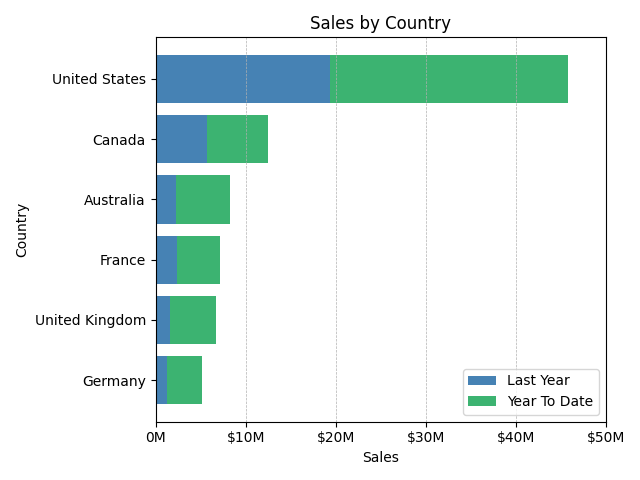
1. What are the regional sales in the best-performing country?
2. What is the relationship between annual leave taken and bonus?
3. What is the relationship between Country and Revenue?
4. What is the relationship between sick leave and Job Title?
5. What is the relationship between store trading duration and revenue?
6. What is the relationship between the size of the stores, number of employees and revenue?

**What are the Regional sales in the best-performing Country?**

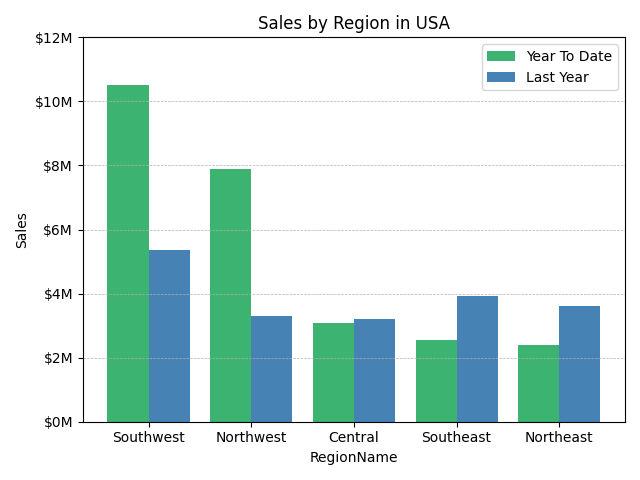
To find this result, we first looked at the sales data from the Sales.SalesTerritory table in the database, which had this year’s and last years’ sales data. As we were unsure of its reliability, we also cross-referenced this by totalling the orders in the database (Sales.SalesOrderHeader) and found similar results:

|  |  |  |
| --- | --- | --- |
| **Territory** | **YTD from Sales.Territory Table** | **Past Year’s Orders Totalled** |
| Southwest | $10,510,853 | $10,752,323 |
| Northwest | $7,887,186 | $8,327,448 |
| Canada | $6,771,829 | $7,014,746 |
| Australia | $5,977815 | $6,319,248 |
| United Kingdom | $5,012,905 | $5,294,242 |
| France | $4,772,398 | $5,179,241 |
| Germany | $3,805,202 | $3,928,539 |
| Central | $3,072,175 | $3,218,890 |
| Southeast | $2,538,667 | $2,661,896 |
| Northeast | $2,402,177 | $2,492,476 |

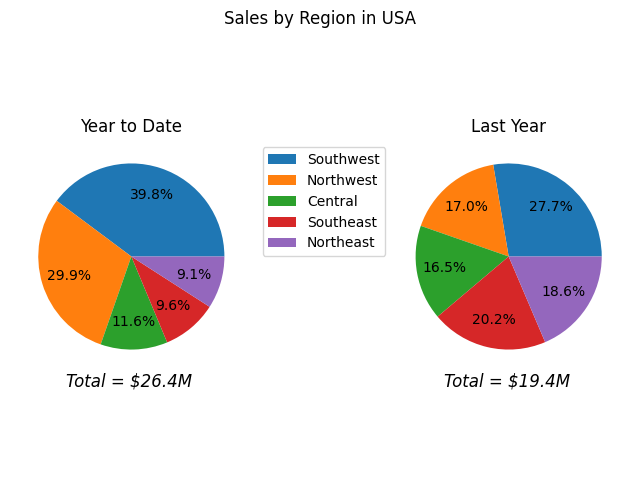
Out of the six countries included in the data, namely United States, Canada, France, Australia, United Kingdom and Germany, the United States was shown to be the best performing by a large margin, with last year’s sales adding to $19,402,505 and this year at $26,411,060, totalling to $45,813,565 over both years.



Within this country, we broke the sales down by region to discover South West was the highest-grossing region over the two years ($10,510,853.87), commanding over a third of the total sales in the year to date, followed by Northwest ($7,887,186.79), Central ($3,072,175.12), South East ($2,538,667.25) and North East ($2,538,667.25).

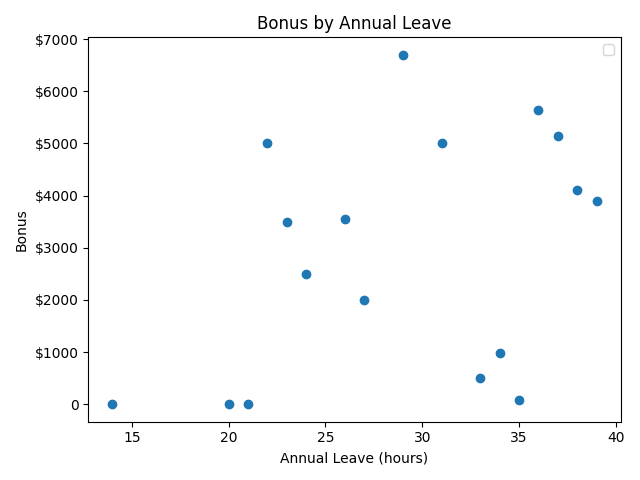


When comparing the two years and inspecting the data further, we can see a relatively large increase in revenue and revenue share from the South West and North West Regions this year compared to last, with the South West’s revenue share increasing from 27.7% to 39.8% and the North West increasing from 17.0% to 29.9%. All while we see a reduced share of the remaining three regions, with Central’s share falling from 16.5% to 11.6%, South East from 20.2% to 9.6% and the North East’s share falling from 18.6% to 9.1%.



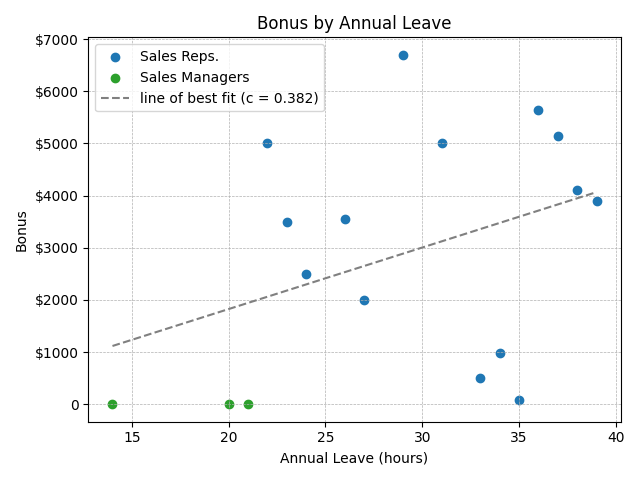
**What is the relationship between Annual Leave taken and Bonus?**

For these results, we extracted the annual leave in hours from the table of all employees and bonus data from the sales department, as they had the only visible data on bonuses. This meant that there was only bonus data for 17 sales employees and this is typically not enough raw data to determine a relationship. The resulting scatter plot is therefore not surprising, with there seeming to be no correlation on initial inspection.



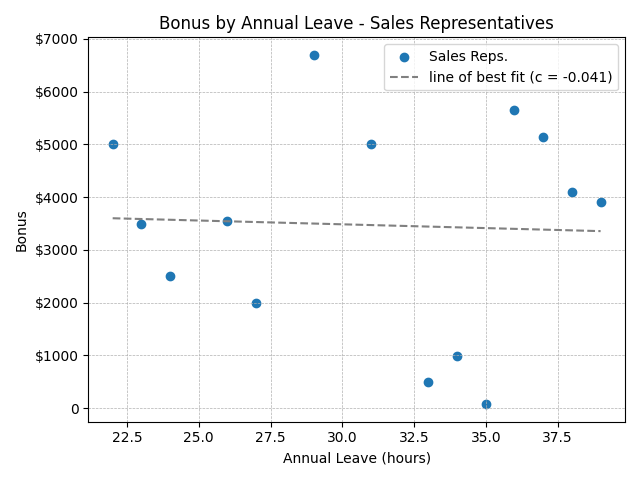
However, when calculating the correlation coefficient (the specific measure that quantifies the strength of the linear relationship between two variables) for annual leave and bonus, we found there to be a moderate positive correlation between the two at 0.3821(0.1–0.3 = weak, 0.3–0.7= moderate, 0.7–1.0 = strong correlation).

This can be visualised by drawing a line of best fit through the data:



In spite of this, it is worth noting that the P-value of 0.13 is greater than 0.1, which infers that there is not enough sufficient statistical evidence to conclude that there is a significant linear relationship between annual leave and bonus.

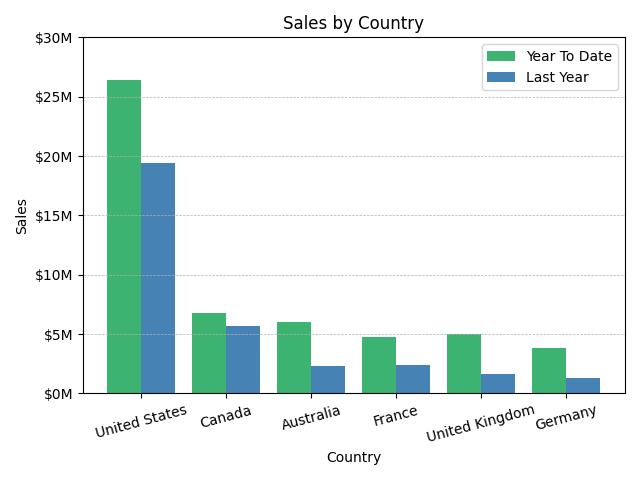
Additionally, out of the 17 employees, only the 14 sales representatives received a bonus while the 3 managers did not receive a bonus. We would need more context on the data for the sales managers as we are unsure if they actually do not receive a bonus, or whether this is incorrect data.

****

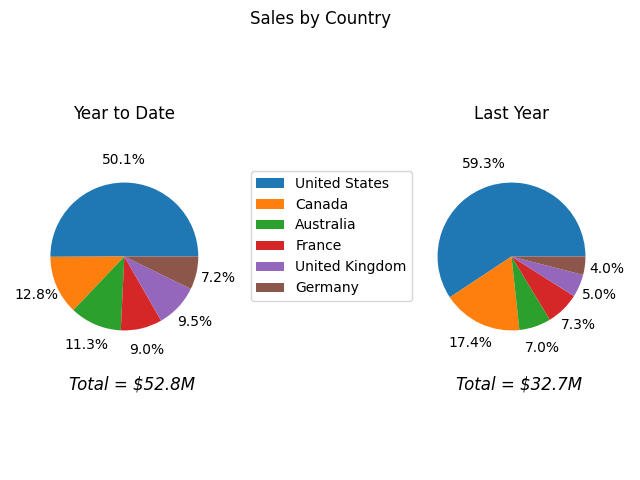
It would be better to determine a relationship for the sales representatives only, however, there is only data provided for 14 sales representatives, so we would need Giant Bicycles to provide us with the complete data for all employees or other departments in order to better determine a relationship.

**What is the relationship between Country and Revenue?**

In order to visualise the relationship between country and revenue, we first found the six countries that Giant Bicycles targets in terms of traffic. From the graph below, we can see that the United States was by far the best-performing country with revenue totalling to almost $46.8m over the last 2 years. In fact, sales have taken a jump from $19.4m last year to over $26.4m this year to date. Canada, Australia and the European Countries are responsible for a much lower proportion of the total revenue, with Canada only slightly improving on sales from last year at $5.7m to $6.8m. It is clear to see from the graph that Giant Bicycles main target market for sales/revenue is North America.

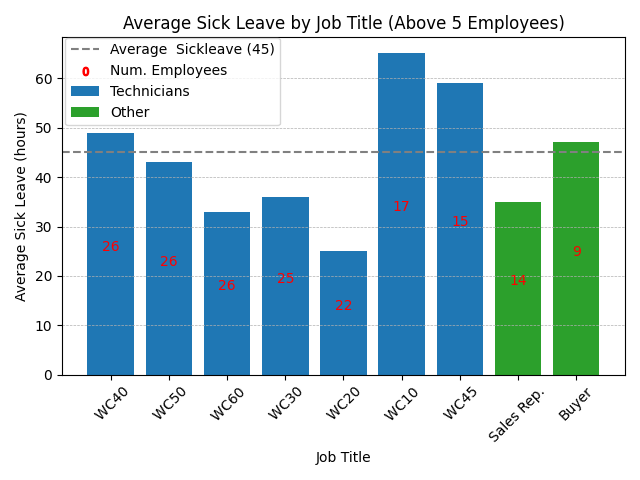


However, it is interesting to see that the markets for Australia, France, UK and Germany have approximately doubled in revenue from last year. This suggests that there is potential for Giant Bicycles to successfully expand their operations in Australia and Europe to maybe break the status quo in North America. We can see Australia is responsible for $6m in revenue this year to date, and at this current trajectory, will overtake the Canadian market by the end of next year. This is supported by the pie chart, showing that Canada’s market has reduced from 17.4% to 12.8%, while the Australian and European markets have grown in the share of revenue from last year by 2-3%.



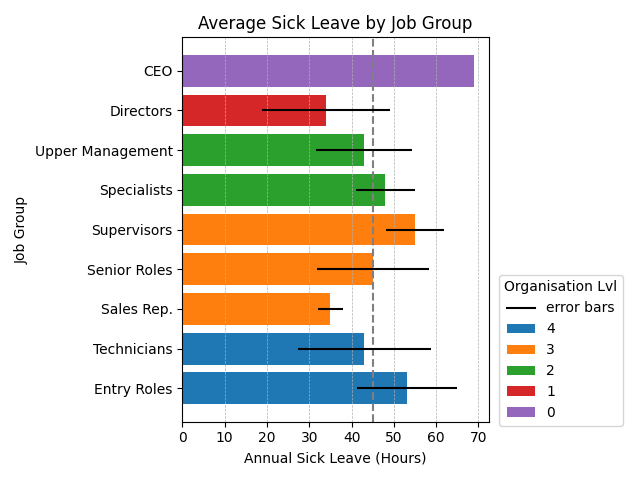
**What is the relationship between Sick Leave and Job Title?**

The graph below displays the average sick leave taken by employees grouped by job title, compared to the average of 45, for job titles with over 5 employees. Unfortunately as there were 290 employees with 67 unique job titles, it is not possible to list many of the job titles as the bar chart would become far too congested with 67 different bars.



To try to encompass the missing data, the organisation level column was utilised in order to group the data together. Level 1 represents employees at the top of the company hierarchy, such as directors or the CFO, while Level 4 represents roles at the bottom of the hierarchy, such as entry-level roles. Additionally, the 290+ employees with job titles were grouped by a new naming system called “job group” to reduce the data points on the x-axis.

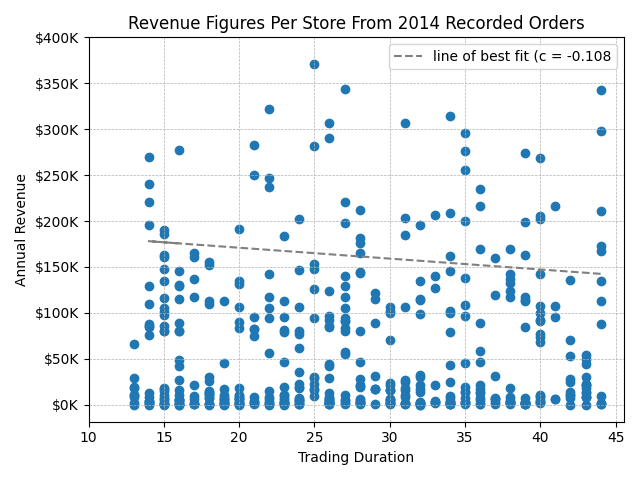
From the graph below, we can see that there does not seem to be much of a trend of sick leave changing as employees move up the hierarchy, but we can see that different job groups do seem to take different amounts of sick leave. For example, Directors and Sales Representatives took 34 and 35 average sick leave hours, compared to the overall average of 45. While entry and senior roles took 53 and 55 hours respectively on average. The error bars show the standard deviation of sick leave taken within each group. The CEO was an outlier at 69 hours and is listed in their own category as they did not have a given Organisation Level.



**What is the relationship between Store Trading Duration and Revenue?**

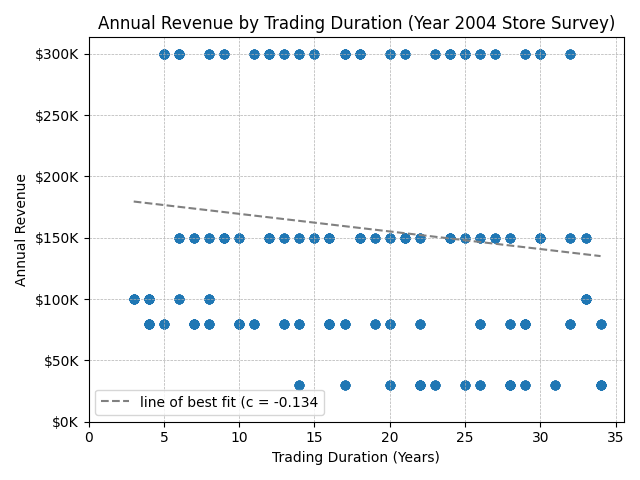
With store trading duration and revenue, we came across a few issues as we had to get the data for this question from an online store demographic survey, which we couldn’t verify the accuracy of. Firstly, the data seems to be imprecise as only 5 unique figures were given in the revenue columns: $300,000, $150,000, $100,000, $80,000 and $30,000, with all 701 stores split into these categories. Secondly, ‘Sales’ figures were always the revenue figure multiplied by 10 in this survey. So bear in mind the accurate revenue values may all be 10x more if the ‘Sales’ figure is actually more correct than ‘Revenue’. But as we are investigating relationships this should not matter.

We tried to cross-reference this with the total orders recorded but unfortunately, many of the orders were not linked with stores. When we plotted stores’ revenue in the last year, there were a lot of data points indicating close to 0 total sales which indicates that stores’ sales were often not being recorded as shown:

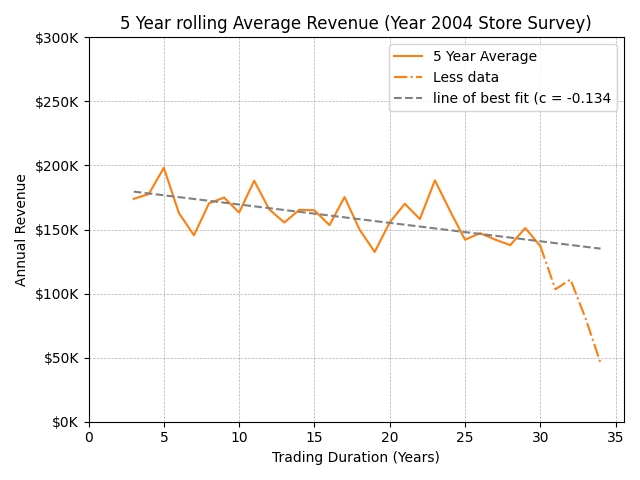


We decided to use the data from the store survey since it was more complete, but the lack of precision in each store’s annual revenue may have affected our ability to make accurate insights from the data.

A scatter plot made from annual revenue data compared to how long a store has been trading for showed a slight negative correlation between the two. And, it is interesting to note that in the last 14 years, the minimum annual revenue for stores has increased to $80k. The minimum annual revenue was $30k prior to then. This could be due to the fact that Giant Bicycles stopped making smaller sized stores within the last 2 decades. The impact that store size has on revenue will be clarified in the next section.



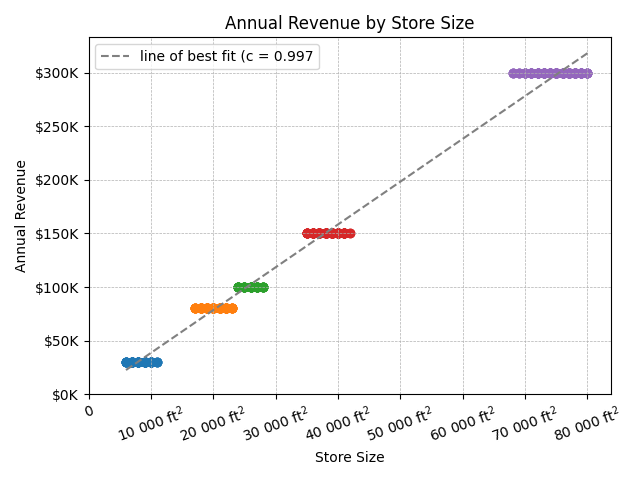
However when looking at a graph showing the 5 year rolling average revenue for stores opened at each trading duration a clearer trend emerges.



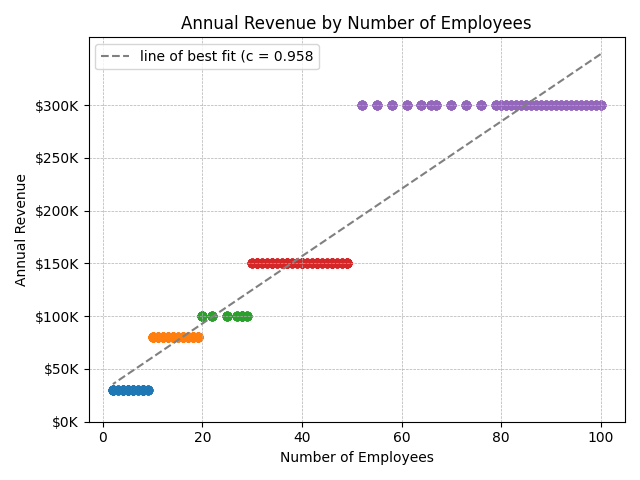
This shows a fairly clear negative relationship between trading duration and revenue, particularly very highest duration stores have much lower average revenue, although there was not as much data for this time period so it could be somewhat inaccurate.

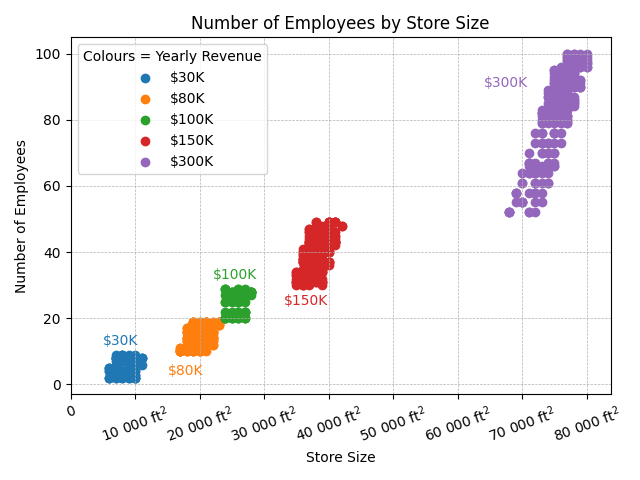
**What is the relationship between the size of the stores, number of employees and revenue**

For this question we used the same store data from question 5, from the online store demographic survey. Again this gave us unusual data, due to the revenue figures being grouped together, but we thought it would be better than the incomplete data from totalling the year’s sales orders. Also once again the revenue figures maybe 10x more than stated, but since we are investigating relationships this will not make a difference.

We can observe a very strong positive relationship between the size of store and revenue, where we can see that stores of around 10,000ft2 have a revenue of $30,000, stores at around 20,0000ft2 have $80,000 dollars, stores at around 30,000ft2 have $100,0000, stores at around 40,000ft2 have $150,000 and stores of around 70,000ft2 have a revenue of $300,000. While these are not necessarily proportional increases, there is a clear trend showing that the larger the store, the larger the revenue.

The number of employees also has a positive relationship with respect to annual revenue. This is to be expected, as the more sales being made, the more staff we expect would be needed to run the store as shown by the graph below.



From the graph above and isolating employees and revenue, it would be understandable to assume that number of employees has a direct positive correlation with revenue creation, however, when put into context with store size, this is where we can see the full story: All three variables are linked, as the number of employees is determined by the store size, which in turn positively impacts the revenue. Stores also seem to be grouped together into 5 clear categories. .****

**Conclusion**

In summary, we have found the following:

1. We found that in the US, the South West region grossed the most, followed by North West. It was also observed that there was a significant upturn in sales from last year to this year in the South West and North West region while observing a fall in the remaining three.
2. On initial inspection, there seemed to be a moderate positive correlation of 0.3821 between annual leave taken and bonus. But we also found this to be not statistically significant. We would need more data to better determine whether there is a correlation or not.
3. We have found that the US is the best-performing country with revenue totalling $46.8m over the last 2 years and the North American market making up 62.9% of total revenue for Adventure Works. However, this has been reduced from 76.7% the previous year, largely due to the fact that the Australian and European markets have doubled in revenue and thus taken up more revenue share. Overall, this suggests that Adventure Works could successfully expand its operations in Australia and Europe.
4. There is no real relationship or trend between average sick leave and hours. Entry-level employees and supervisors tend to take more sick leave, while Directors usually take the lowest amount of sick leave on average.
5. When grouping the data into 5-year averages we started to see some clear positive correlation, with our results stating that stores tend to return greater revenue than those opened 5 years prior.
6. There seems to be a very clear and strong positive correlation between store size, number of employees and annual revenue. The larger the store sizes, leads to more employees, and therefore, more annual revenue.

We should stress again that the data used for questions 5 and 6 was not ideal, so we would recommend that Giant Bicycles tries to ensure that all future orders are entered into the database with a corresponding store, ensure accurate figures for annual revenue, to make it possible to gain a better insight about how stores are performing.

We would also need bonus data for more employees or other departments to better determine the relationship between annual leave and bonuses.