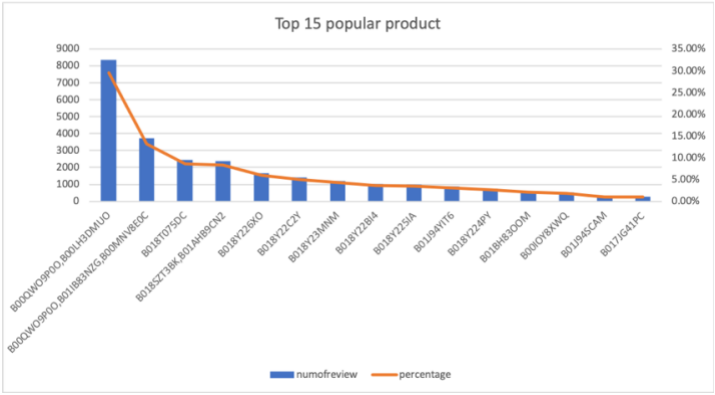
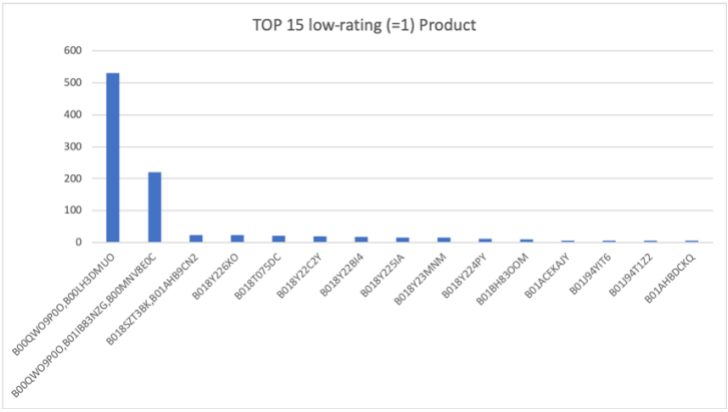
Data Analysis on Amazon Product (SQL, EXCEL, PYTHON)

1# The most popular products in Amazon.com. And the products with most of low rating.

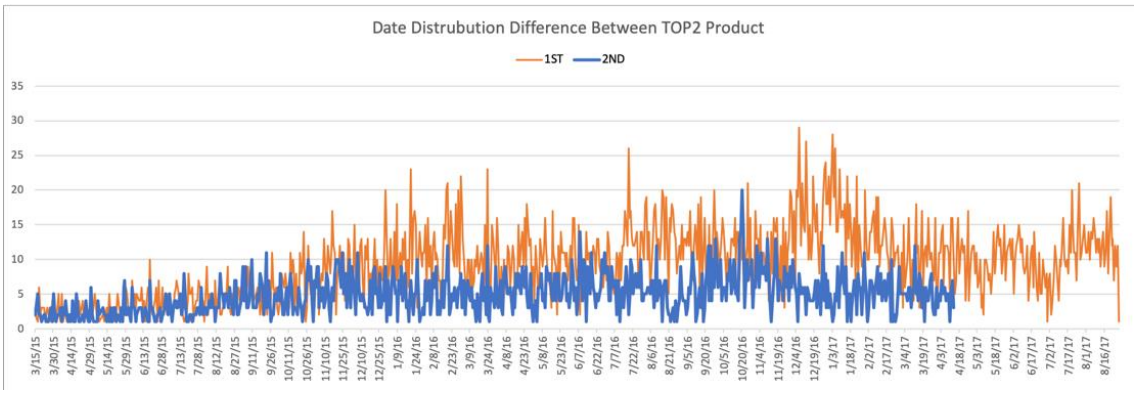
The top 15 most popular products are shown below:





Based on the result, we can conclude that:

- The product with #'B00QWO9P0O,B00LH3DMUO' has 8343 reviews (29.45%).
- The product with #'B00QWO9P0O,B01IB83NZG,B00MNV8E0C' has 3728 reviews (13.16%).
- The number of reviews difference among products is significant.
- Top 2 popular and Top 2 low rating products are same. This is consistent with the regular market rules.



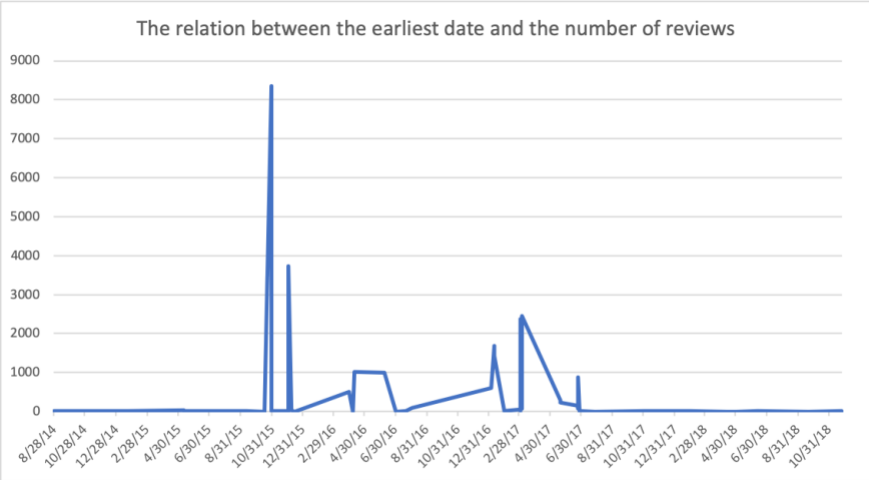
Also, we can see the 1st product is more than 2nd product. In order to explore the reason of this, we analysis the date distribution between two products. The following figure show that:

- Two products are added at similar date, but 1st product has longer product sell duration.
- The 1st product keeps a higher sales volume.

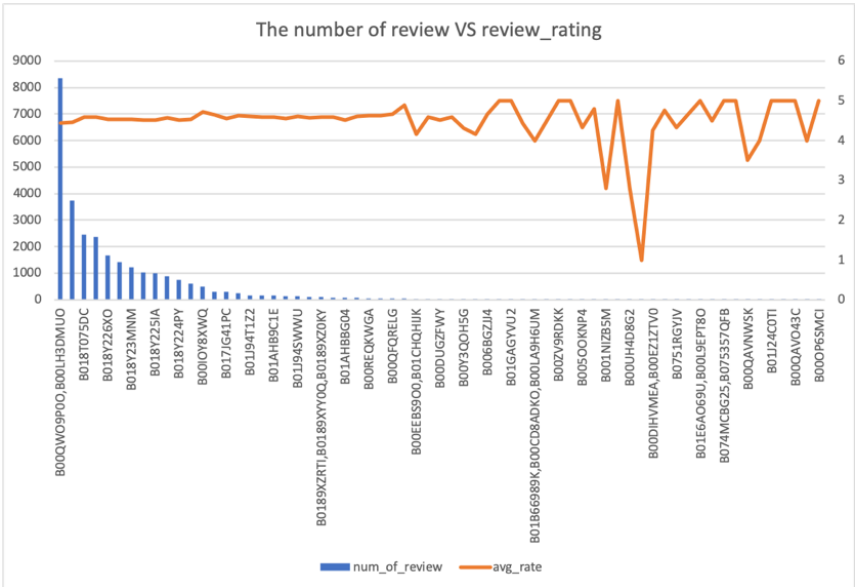
2# The relation between the number of reviews and the product-added date.

From the graph, we can see that:

- There is no linear (positive/negative) relationship between the number of reviews and the product-added date.



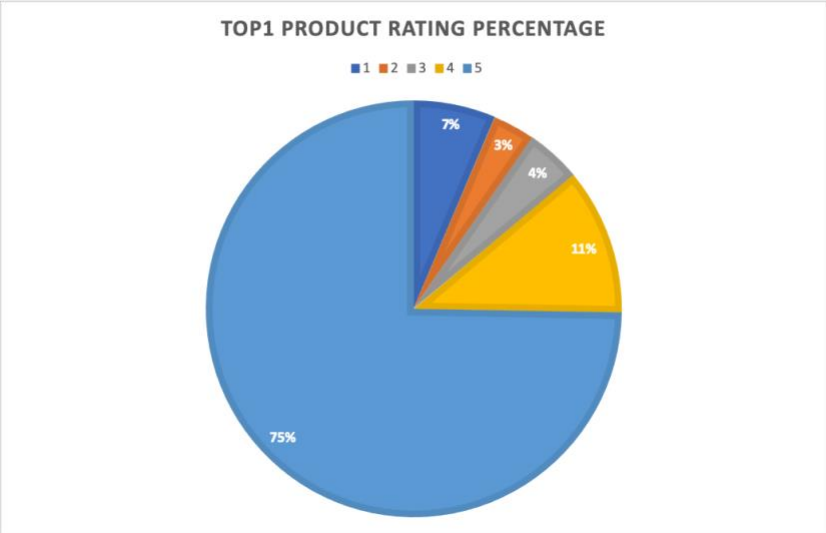
3# The average rating of each product. And the relation between number of reviews and rating.

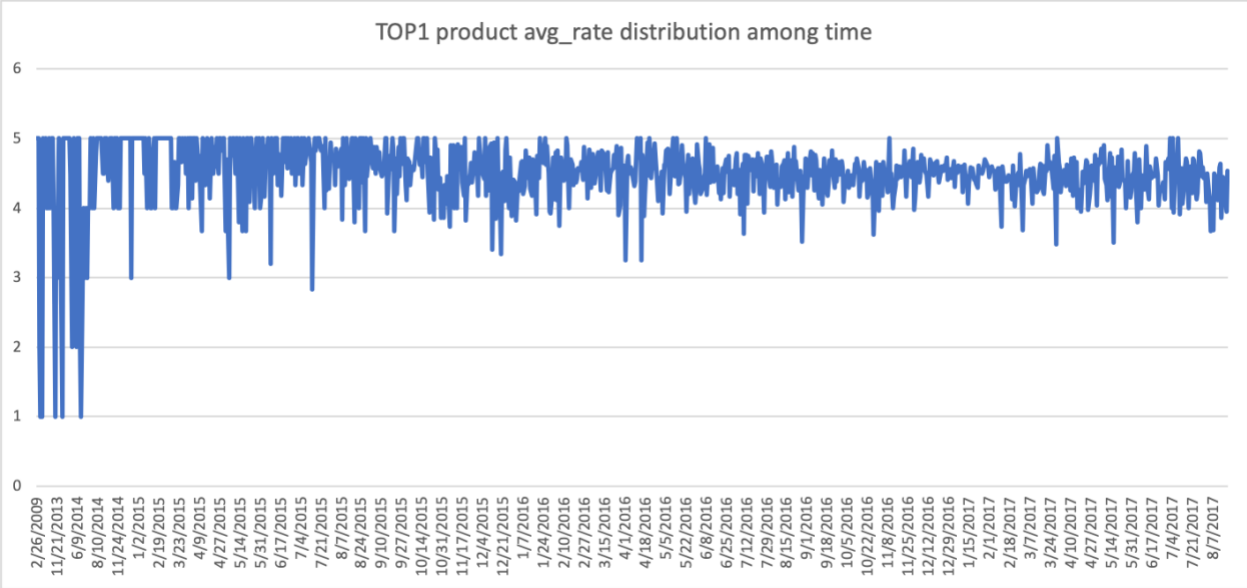


- The average rating of most products is around 4-5.

- There is one product, which has a small number of reviews and the lowest rating. The producer can aim at this product, find the reason and improve this.There is no obviously relationship between number of reviews and rating.

4# Deep diving into one specific product, the rating change among different time.



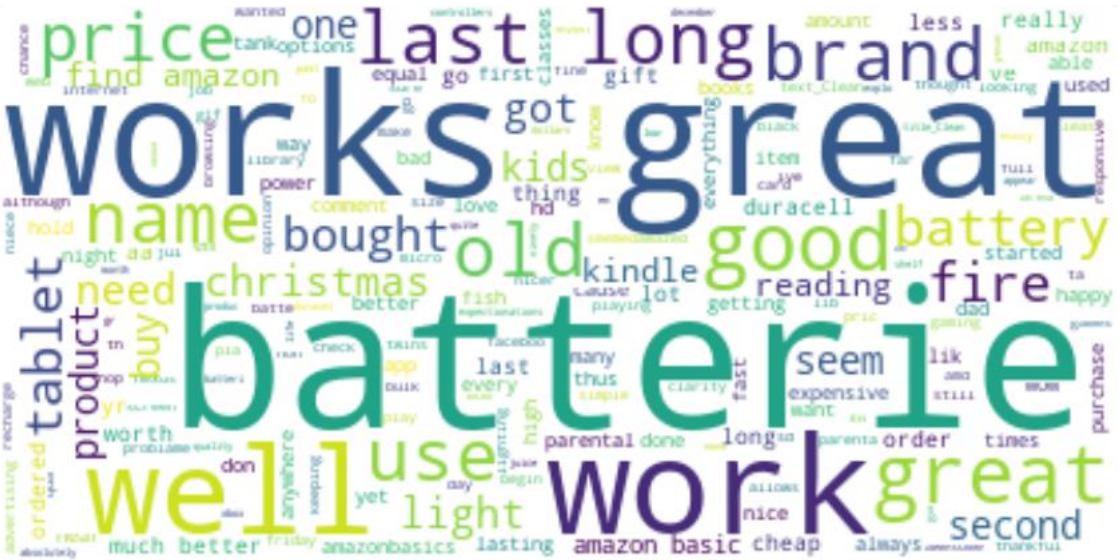


For top 1 product, we can see that:

- Most of the rating of top1 product is 5.
- Most of the low-rating of top1 product is posted at the date the product added and the first three years. After that then the rating increases, it may because the producer improves the quality of the product, or change the price.

#5 WordCloud Based on product review text and title.

#extract review with rate score >=4 as positive review.



#extract review with rate score <=3 as Negative review.

From results, we can see that:

- For positive review, most of the comments seem positive, like "good","great","well".
- For negative review, the number of positive words obviously decrease.
- Also, we can conclude that battery is the most popular product.