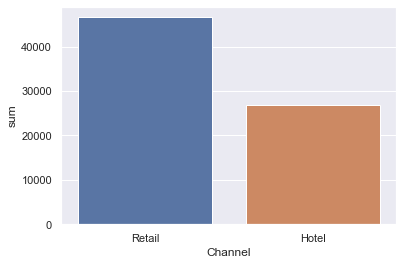
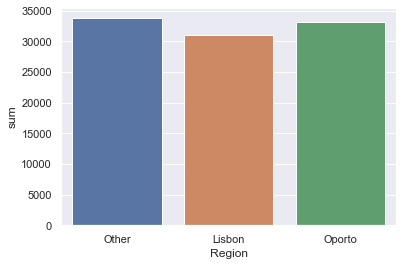
* 1. **Use methods of descriptive statistics to summarize data. Which Region and which Channel seems to spend more? Which Region and which Channel seems to spend less?**

**We can see the visualization for sum, which is the sum of all the columns (Fresh, Milk, Grocery, Frozen, Detergents Paper, Delicatessen)**

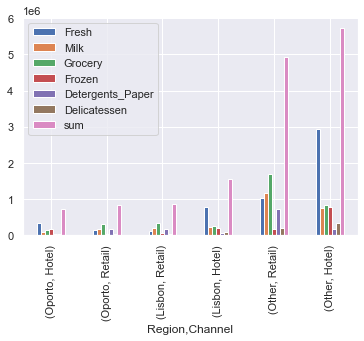
**Retail channel tends to spend more and Hotel channel tends to spend less.**

****

**Other region spends the most and Lisbon region spends the least. This data we got but using sum column. Now, we will inspect individual columns to see and irregularities for specific region or channel.**

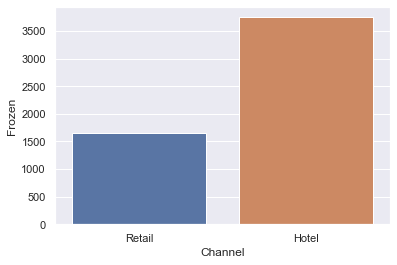
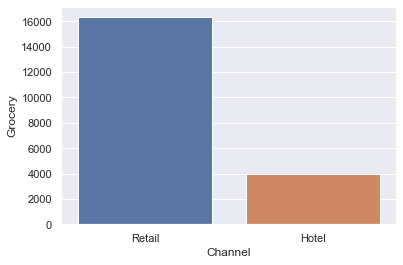
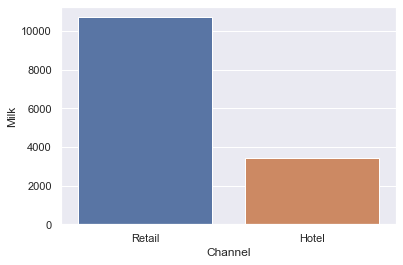
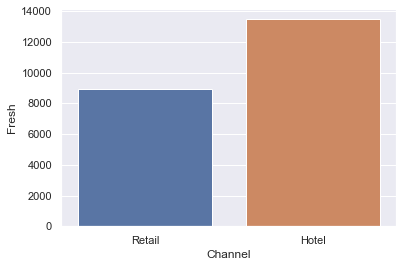
****

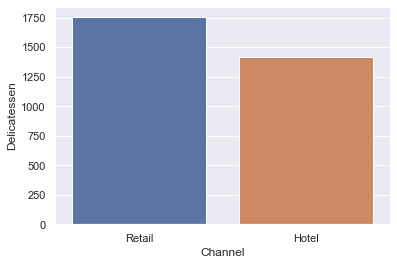
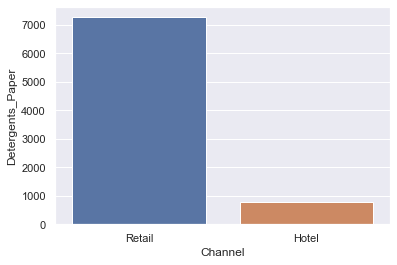
**The bar graph for individual region and channel have also be added. If we are looking for the combination of channel and region which has the max output.**

****

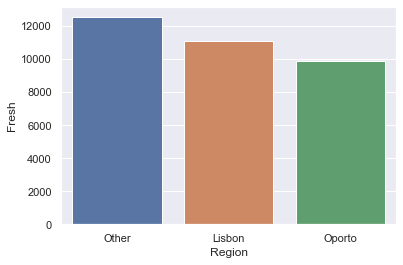
**We can see that some region have massive produce from a single channel, but to have a more general approach we, have considered the sum column to get the most general data.**

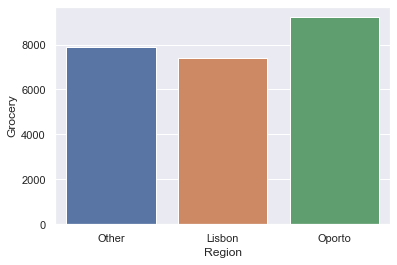
* 1. **There are 6 different varieties of items are considered. Do all varieties show similar behaviour across Region and Channel?  Provide justification for your answer**

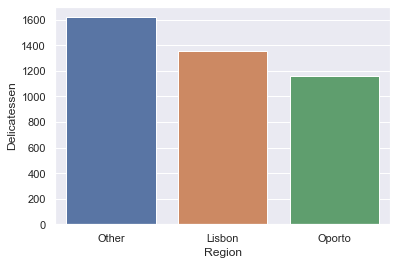
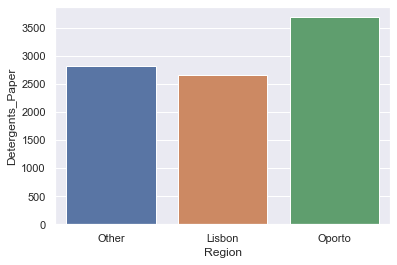
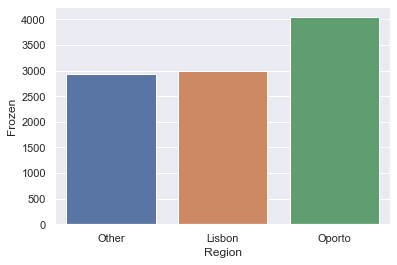
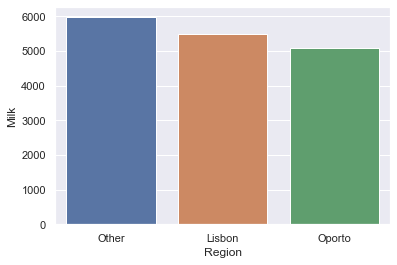
****

****

**Now we do it across regions.**

****

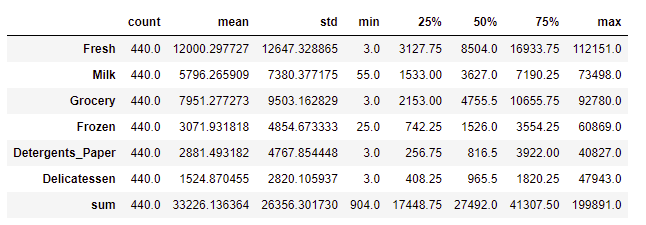
****

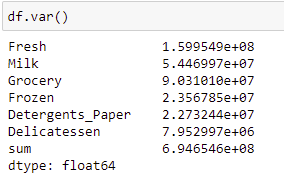
****

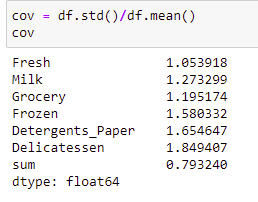
**For all varieties across region the data says that the distribution is not equal but is definitely similar with the products distributed evenly across the three regions more or less.**

**Where as for varieties across Channels the data is very inconsistent. In some varieties the difference between hotel and retail is too drastic.**

* 1. **On the basis of a descriptive measure of variability, which item shows the most inconsistent behaviour? Which items show the least inconsistent behaviour?**



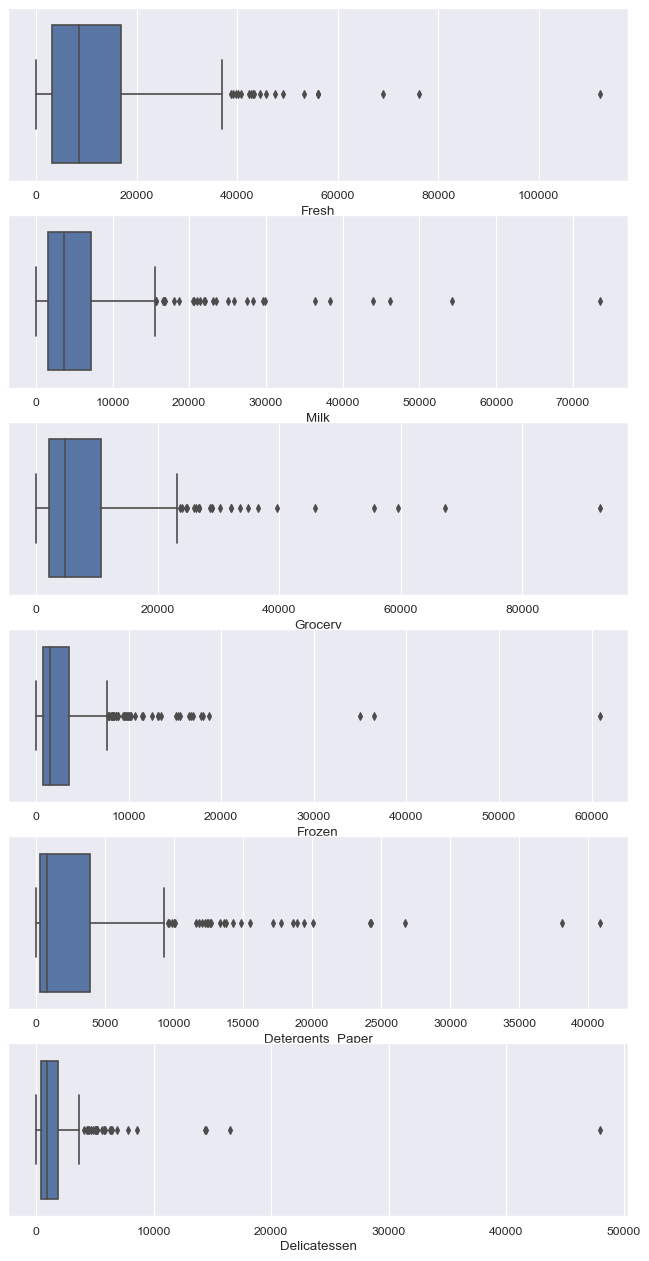




**We found that variance was not the best parameter to compare the data. So, we normalized the standard deviation using mean. According to our result the most inconsistent is Fresh and the least inconsistent is Delicatessen.**

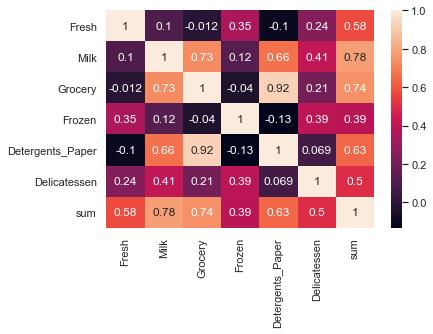
* 1. **Are there any outliers in the data?**

**Yes, there are outliers in the data.**

****

* 1. **On the basis of your analysis, what are your recommendations for the business? How can your analysis help the business to solve its problem? Answer from the business perspective**

1. **We can definitely help the business by working on the inconsistencies in the data, for example when we saw varieties across Channels was very inconsistent. We can analyse the reason behind this, does this have something to do with the culture or the season or geography of the place. According to this step can be taken.**
2. **We can identify the correlation between varieties and use it to our advantage by using buy 4 and get 1 free type of schemes leveraging the popular products.**
3. **We can also change the layout of our stores keeping the popular product in the far side so that the customer will have to walk a certain distance. During which, he might buy something he/she/they saw on the fly.**

****

****