# Descriptive Data Analysis Notes

Before starting with descriptive data analysis, we contacted the client to solve the previous doubt about the field Senior. They told us that the Senior category should include 65 year old subscribers. Therefore, we applied the following conditional function to that column:

=SI([@Age]>65;"Yes";"No")

Once we had completed data cleaning and transformation, we copied the excel file into a new folder called Data Analysis and then we renamed that file to “amazon\_churn\_descriptive\_analysis.xlsx”. Descriptive data analysis was performed on that file.

For this purpose, we used the Data Analysis tool that is already integrated in Excel, which enabled us to obtain a quick summary of the most relevant descriptive statistics: mean, median, mode, standard deviation, variance, minimum, maximum, range, skewness coefficient and kurtosis coefficient.

## Descriptive analysis: Insights on numeric columns

Insights are highlighted in yellow. Further analysis in blue.

* **Account Length (in months):** median (28 months) and mean (32.31 months) are similar, which suggests an equilibrated distribution of the accounts antiquities. On the other hand, a mode of 1 month indicates the majority of subscribers are new clients. However data dispersion is notorious: we have very new clients but also very old clients (with an account length up to 77 months (almost 6.5 years). This is confirmed by the big standard deviation and variance. Account Length data reveals the existence of two groups of customers: a significant group of new customers characterised by high rotation rates and an extended group of loyal customers with long-term retention rates. This variety of profiles suggest the need to apply different marketing strategies for each segment.
* **Customer Service Calls:** the average number of calls is less than 1, meaning the majority of subscribers did never called. This is confirmed by the mode (0) and the median (0) which suggest that more than half of the subscribers never called. On the other hand, data is really dispersed: there is a minority of customers that have called a lot of times, versus a majority of clients who are not calling at all. This is supported by the range from 0 to 5 and positive skewness. Clients who did multiple calls may require special attention and the reason for churn may be associated with these issues.
* **Average Monthly GB Download:** the average number of downloaded GB per month is approximately 6.7 GB. The 50% of the subscribers download 5GB per month or less. Moreover, the majority of the subscribers never downloaded films or series from the service, as the mode is equal to 0. Probably the streaming option is the favourite. On the other hand, data is really dispersed: there are customers that have downloaded a lot of GB per month versus customers with no downloads at all, which can be deduced from the range, standard deviation, variance and positive skewness. Data indicates the majority of subscribers prefer the streaming option while there is a significant group with a preference for downloading content. Segmented groups may need diversified marketing strategies and service plans.
* **Extra Data Charges:** the average extra data charges is almost 3.4 USD. Moreover, the majority of the subscribers have not been charged extra as the mode is equal to 0 USD. Same for the median. Most probably, this variable is positively correlated with the above (average monthly GB download): the more the GB downloaded, the more the extra data charge. Check this. On the other hand, data is really dispersed and grouped: there is a minority of customers that had been charged great quantities due to their particular need for big downloads (they may need special attention) versus a majority of customers who strictly control data usage and have no charges at all. This is confirmed by the range value (99) and extremely positive kurtosis.
* **Age:** the average age of the subscribers is equal to 47.53 years old and is mostly identical to the median (47), suggesting an equilibrated distribution, close to normality and, in general, mid-age subscribers. In addition, there is a majority of subscribers who are 29 years old (mode). However, data dispersion is notorious: the standard deviation is approximately 17 years old and the range goes from 19 and up to 85 years old. It is clear that Amazon Prime Video offers a variety of contents suitable for all ages and that the company plays in a very segmented market, needing to apply very diverse marketing strategies for assuring client retention.
* **Number of customers in Group**: the average number of customers in Group is less than 1, meaning the majority of customers are enrolled in an Individual Plan, which is confirmed by the mode and the median values (0). On the other hand, data dispersion is a fact: a significant group of customers who are enrolled in a Family Plan coexist with a majority of customers in Individual plans. This is confirmed by the standard deviation, variance, range (6), positive kurtosis and positive skewness. Both types of plans are suitable for Amazon Prime Video’s customers.
* **Monthly Charge (USD):** the average monthly charge is equal to 30.90 USD/month. The median is very similar: 31 USD/month, indicating an equilibrated distribution. On the other hand, the majority of subscribers pay a charge of 10 USD/month, which reflects that, in general, clients prefer more economic plans. However, the standard deviation and the very wide range (78) suggest that data dispersion is notorious, indicating a huge variety of plans, which can be strategic to reach more market segments.
* **Total charges (USD):** the average total charges is equal to 1084.78 USD. The median is much lower: 647 USD, meaning that the 50% of subscribers paid a total of 647 USD or less, which is relatively low. The majority of subscribers pay a charge of 10 USD/month. On the other hand, data dispersion is extreme: standard deviation (1129 USD) and range (5568 USD) have very significant values, meaning that total charges are really diversified, which may indicate a need to segment marketing strategies and customer services to cover necessities of all this groups.
* **Average Monthly Expenses (USD):** the average of the average monthly expenses per subscriber is equal to 30.98 USD. The median is very similar: 30.66 USD, meaning it is an equilibrated distribution. However, data dispersion is notorious: the standard deviation is approximately 16.84 USD. This idea is reinforced when we observe the range of monthly charges (80.67 USD). All in all, the variety of expenditures, with values from 4 USD to 80 USD distributed in an equilibrated manner, show a clearly segmented market where different marketing strategies need to be applied for customer retention.
* **Number of Complaints or Support:** A median (5) almost identical to the mean (5.01) and skewness coefficient close to 0 indicate that clients interaction with the customer service is more or less equilibrated and more or less similar, without significant segmentations. However, data is dispersed: it reveals there are customers registering up to 10 complaints or support. This means, customers who complain or request support often do so multiple times, potentially indicating unresolved issues after the first interaction. This is something that requires special attention. A correlation analysis between churn and the number of complaints or support requests could help determine whether frequent complaints contribute to customer attrition.