Module 4 Assignment

1. State the problem described by the client & desired outcomes.

The Chief Marketing Officer (CMO) of a Fitness chain has the following requirements:

* Identify the root cause for high churn rate across the fitness chain
* Why are subscribers increasing after holidays and the rate disenrollment drastically increasing the rest of the year?
* Understand the demographic of the customers leaving and joining
* Should a new marketing strategy be used? A new product launched? New payment method? New demographic be targeted?
* What would be the economic impact if a new a strategy is to be implemented to retain customers?

The desired outcomes by the client are as follows:

* Identify the root cause for high churn rate across the fitness chain.
* Explain the revenues pre and post holidays and how the revenue is performing in the current marketplace.
* Identify reason for increase in subscriptions post holidays and loss of subscriptions the rest of the year.
* Identify the demographic to be targeted.
* Should current strategy being implemented be improved or a new strategy is required to lower churn rate? The economic impact of implementing this strategy.

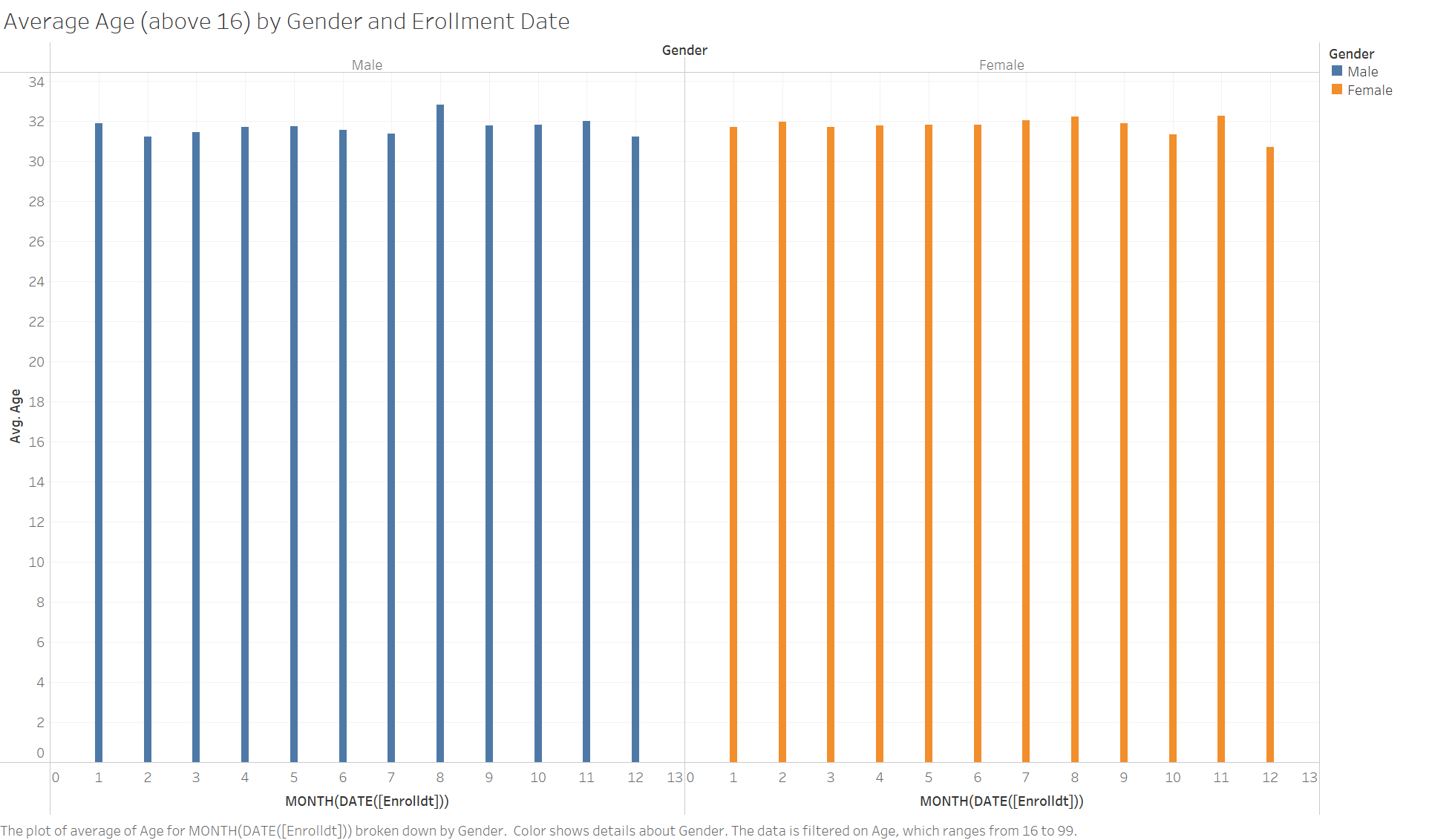
1. Describe your hypothesis identifying the business drivers (influences).

Based on the data given by the CMO and my research on fitness chains in the market, I identified the following business drivers for the current Fitness Chain:

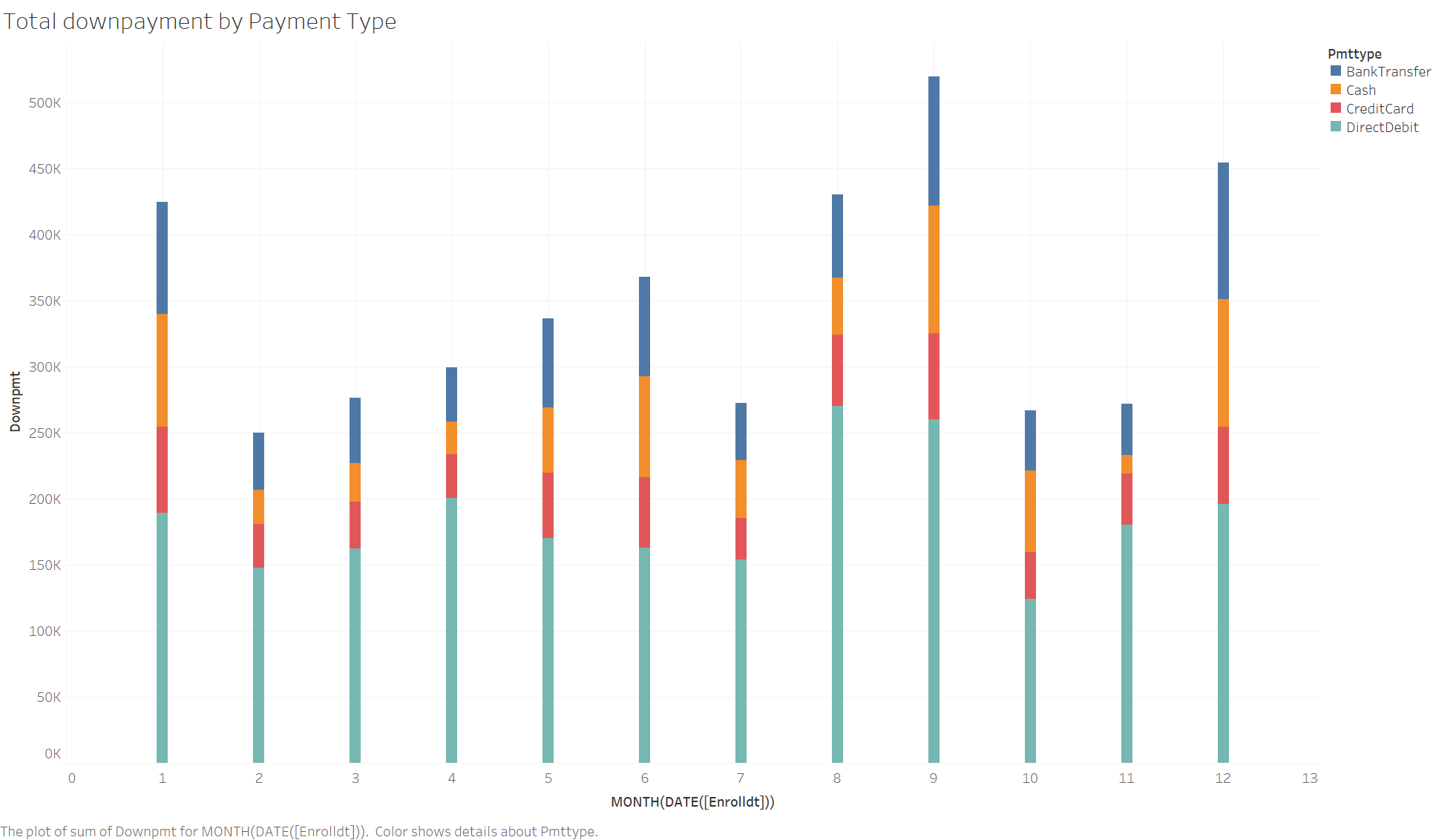
|  |  |
| --- | --- |
| INTERNAL FACTORS | EXTERNAL FACTORS |
| Operations like workflow, staff, equipment | Customer satisfaction |
| Location of fitness centres | Competitor’s position in the market |
| Marketing strategy | Economy |
| Sales rate | Population |
| Churn rate | Consumer habits |
| Partnerships |  |

Using the data given:

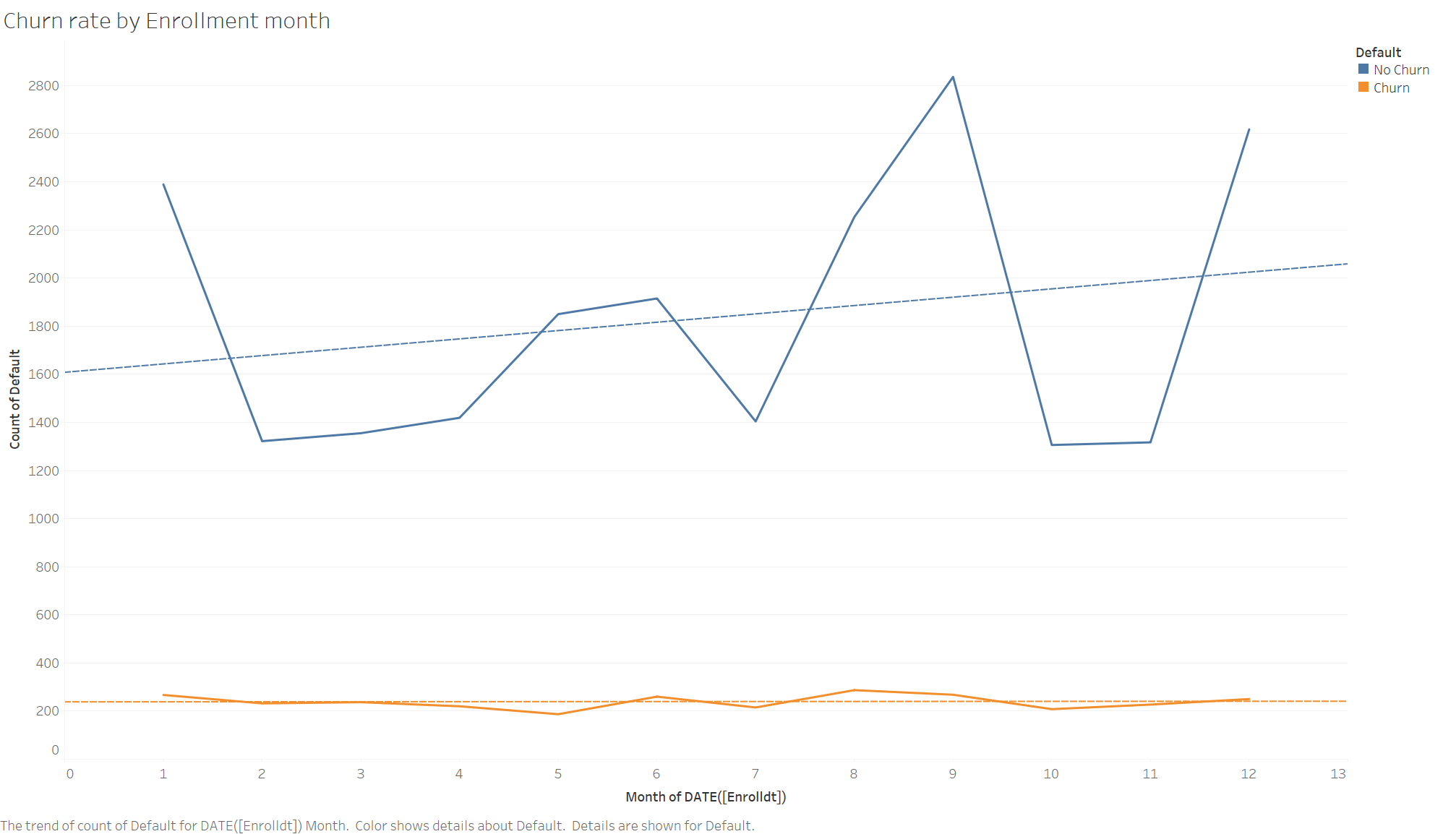
1. I created a histogram of average age and gender enrolling over the course of the year 2019 and beginning of 2020 assuming the date is stored in the format (yyyymmdd). From this I identified that gender is not playing much of a role in the enrolment and the average age for those joining the fitness centres is around mid to late twenties. More information on the ethnicity or location is needed to identify the demographic currently enrolling.



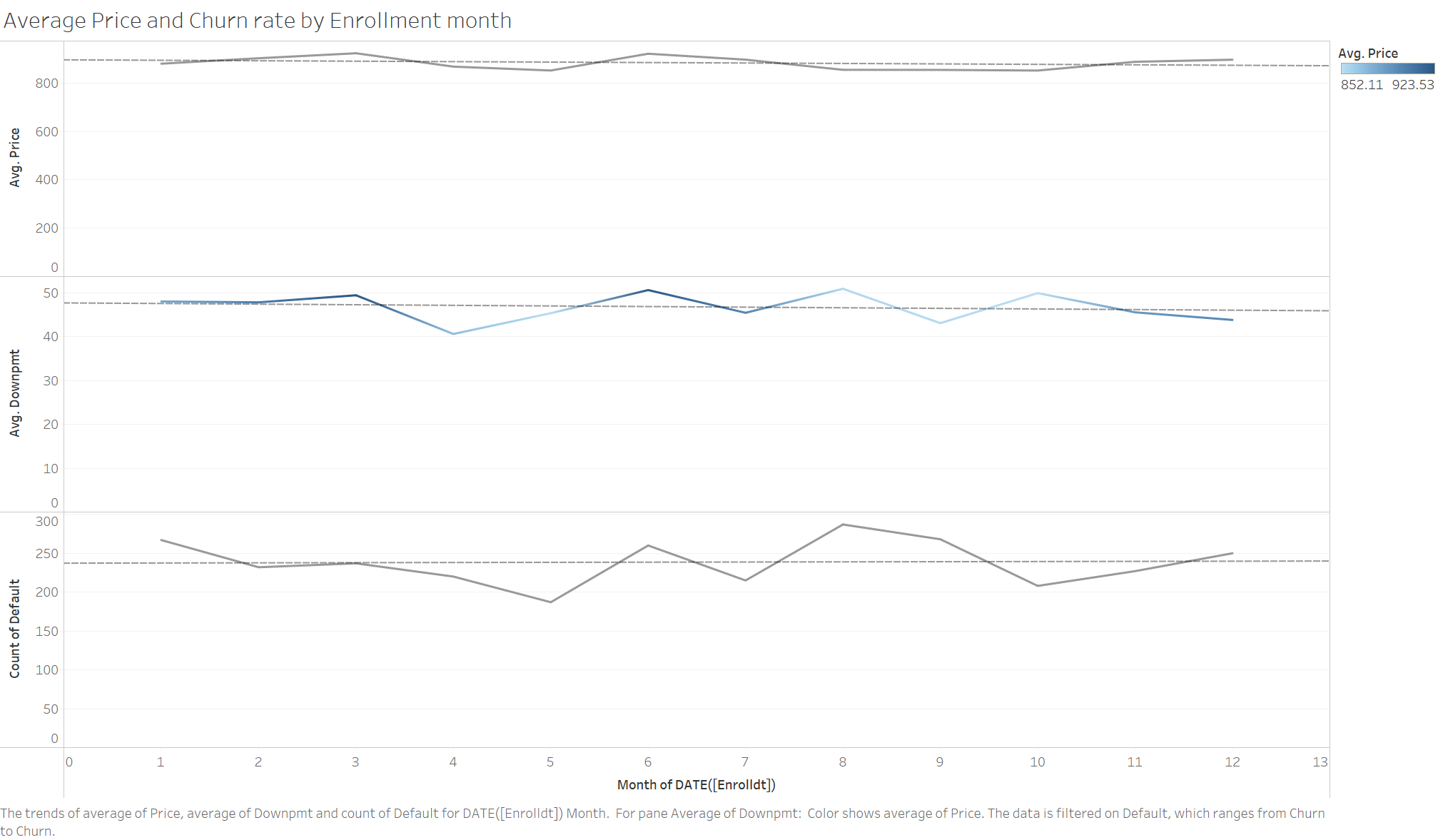
1. I created a Bar chart to identify the maximum down payment received in the year across the various payment types. Based on this, it can be seen that during the months of August, January and December, the down payment is the highest indicating that before or during holiday season more people are signing up. This is similar to the information given by the CMO and can be analysed more by getting cleaner data to understand. While payment type used the most is ‘Direct Debit’ followed by Bank Transfer, Cash and Credit Card, this information is ambiguous. More information needs to be gathered on what is meant by Direct Debit as it generally is an auto payment method which may not be accurate in case of down payment. Bank Transfer/Credit Card may mean that it is online purchase for a subscription/App/physical product like cycles. As mentioned, information on products offered needs to be gathered.



1. The major question asked is what is the churn rate looking like across the year for the fitness centre. I created a line graph of this along with a trend line. While it does show that customers are consistently leaving it also shows that during months of August, January and December, the number of customers staying loyal increases. This is somewhat ambiguous as the methodology for churn rate collection is not provided in detail.



1. As seen below, as the down payment price increased, the number of defaulters increased in that month.



1. Identify data issues that might influence model selection.
2. The methodology of collecting data might not be refined. For example, in the given data, doesn’t collect date in any particular format like yyyy/mm/dd or mm-dd-yyyy. Instead it is a concatenation of year, month and date which makes it difficult to identify which is the month or date as different places follow different formats.
3. Incomplete/Inaccurate data – The data collected contains many NULL values or consists of incomplete or inaccurate information. For example, the given data has age as 0 for many entries which makes the data useless for use. It also doesn’t provide more information like locations, names, subscriptions taken up, subscription details, etc.
4. Insufficient/Excessive data – The data collected may be too small or too huge to perform an accurate analysis and model fitting. The model may then under fit or over fit with a huge margin of errors and lack of predicting power.
5. Duplicate data, outliers – Duplicate data needs to be cleaned up to avoid errors in the model while outliers may significantly impact statistics like mean and standard deviation leading to bad prediction or estimation.
6. Identify all factors that might influence model selection process:

Factors that may affect model selection are:

* Data – The data collected, if it is too less or is incomplete or having inconsistencies, then a simpler model may be used to fit the data, however the scope of the model will also reduce.
* Requirement to be answered – Based on the question to be answered, the model is chosen. For example, if the question is to identify the sales rate and margins after the implementation of a new strategy, a predictive model may be used. If a target audience needs to be identified, a clustered model may be used.
* Timeline – Generally, the deadlines for projects are very tight. In this case, the data will have to be quickly analysed, fit to a good model, generate insights ad present to the clients. Keeping these deadlines in mind is crucial as models that may take time cannot be used in such situations.
* Cost for tools and resources – The company to work on the project should have the resources to store and use licensed tools if necessary based on the question to be answered. If the question is simple and the model required can be done in something like Python, then it’s easier.