1. Today is 29 of April of 2021 (the last call\_time on our database from calls/voicemails/customer\_service
2. Peak times are from 7 am to 7pm regardless if it’s weekday or not. ( I actually just saw the email after I finished nearly everything and I found risky, due to time, to change the definition of Peak time)
3. Call rates are applied for the time when the call/voicemail started, even if it goes from off-peak to peak time or vice-versa in the middle of the call
4. A Call/Connection, can only have one rate\_type. If a call is\_international or is\_roaming and also is\_peak, the is\_international/is\_roaming takes precedence.  
   Precedence rule: is\_international>is\_roaming>is\_peak>is\_off\_peak
5. Stakeholders of this project can come from a diverse background, including business people or even former/current DBAs that never worked with denormalized data, with that, it’s important to give a pathway of what would be possible to build, why it’s not a good idea and then what we should build to solve our current business requirements.
6. Brand new star schemas or even changes on current schema might come often and we need to have flexibility and agility to produce new diagrams and collaborate on them remotely, that’s why I chose an online diagramming tool like [Lucidchart](https://www.lucidchart.com/pages/) instead of the OracleDataModeler even though OracleDataModeler would give more details about the model that I was building.
7. Stakeholders don’t mind to have data from a few hours ago on their reports but would mind to have reports that take hours to run so, we can worry less about how long it takes to build the star schema and worry more about the performance it will deliver when using it.
8. Call duration is in seconds and, for charging, is rounded up to the next minute
9. By churn, I understand customers which the end date is smaller than the definition of today (29 of April of 2021 )
10. I’ve realised only on finishing the second section that phone\_number wasn’t a unique identifier for customers (there is one phone number that seems to belong to two different people), that’s why I haven’t committed the same mistake on section C but was late to correct the other sections. In any case, as it just one out of 5000, it shouldn’t impact that much the overall analysis.
11. I got the generations information from here: <https://en.wikipedia.org/wiki/Generation#:~:text=A%20generation%20refers%20to%20all,and%20begin%20to%20have%20children.%22>
12. As I didn’t know exactly which values to plug on the ML algorithm and I didn’t wan’t to jot everything in there and hope for some good result, I just created a notion of “attributes important for the business”. That resulted in a Moneyball thing were at least I’ve proved that the “attributes important for the business” weren’t that important after all.