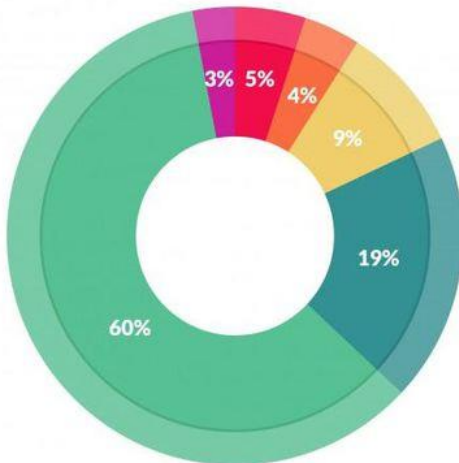


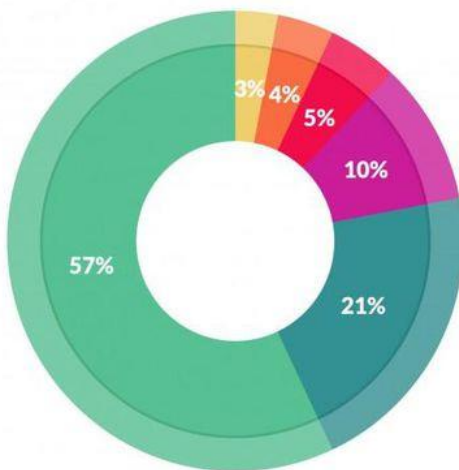
A long Speech

As a Data-Scientist a huge part of your time will be taken up by simply collecting, cleaning and organizing data. I assure you, this portion is the least fun and yet the most important part of our role. You will hate it from the bottom of your hearts and wish for the day when you can finally get to try out your fancy new algorithms. Heck, you will spend hours just waiting for that first chart you get to make. But that's the thing, the biggest asset of a Data Scientist is not all the algorithms you have learned or the tools you have practiced with. It's patience and the ability to find patterns.



What data scientists spend the most time doing

- Building training sets: 3%
- Cleaning and organizing data: 60%
- Collecting data sets: 19%
- Mining data for patterns: 9%
- Refining algorithms: 4%
- Other: 5%



What's the least enjoyable part of data science?

- Building training sets: 10%
- Cleaning and organizing data: 57%
- Collecting data sets: 21%
- Mining data for patterns: 3%
- Refining algorithms: 4%
- Other: 5%

With that in mind, this a sample problem we are facing right now.

Problem Statement

A couple of months back, we were rushed to make a dialing interface live. Thanks to the quick and dirty product launch, someone left the User-Comment field as a free text field. Which means now we have illegible text instead of the beautiful clean data I was expecting.

Using other methods, we were able to extract Event-Type for each event.

But this adversity is also an opportunity. Now we want to now see how well an NLP model can predict Event-Type based on User Comments.

Your submission must include

- 1) A report detailing the steps you took and any results that you generate. Write down all your thoughts or inferences.
- 2) Any codes that you used. You are free to use any language of your preference, although I would very strongly encourage using Python (only if you are familiar with it). The codes should be well structured and commented.
- 3) Predictions for test-data.
- 4) Bonus Points for any other cool insights you find.
- 5) Bigger Bonus Points if you can extract added information at Event Level.