Which topic did you choose to apply the data science methodology to? **(2 marks)**

**Emails**.

**1.Problem related to the topic I selected is as below:**     
Receiving the mails that are not relevant or of no use to the user anymore which they once subcribed while reading some article or news on internet which  eventually lead to unnecessarily filling up the inboxes.  
**2.Phrase the problem as a question to be answered using data**  
  
Usually it happens that while browsing some content on internet we subscribe to some pages thinking that it might help in future as well for some information but in reality after few days we totally forget about that subscription as they are now of no use to user. This leads to unnecessarilly filling up of the user mailbox.  
My question is how can the user should avoid getting irrelavant mails so that their mailbox should not be filled with random messages or notification.

1.ANALYTICAL APPROACH

We can start with deciding a period for which we can analyse all the mail received in that particular period. For example we can consider all the mail received for last 6 month. We can divide all the mails into READ and NONREAD.The READ mail we can leave as it is. We have to see the pattern of UNREAD mail. In that pattern of unread mail we can again check the frequency of mail and the id with which they are associated or received from. Then we will see if any mail coming from same subscribed pages have been read before or not. If not then we can unsubscribe that page because then it means that if we have not read that particular subscribed mail in last six month that means we don’t need their message anymore.

2.DATA REQUIREMENTS/ Email for last six months will be required to analyse the need of the particular subscribed mail.

DATA COLLECTION. We can collect data from different email account to check the no. of mails read or responded by user to that particular.We can also check what percentage of the subscribed page has been responded or not responded.If responded mail is 20 to 30 % of total unread mail then we can leave else we can unsubscribe the page.

DATA UNDERSTANDING and PREPARATION.

we can get the data in csv file. Can use pandas to get them in dataframe. Can name column as read ,unread,and the frequency of mail etc.

DATA PREPARATION. In this section, we will prepare data for the next stage in the data science methodology, which is modeling. This stage involves exploring the data further and making sure that it is in the right format for the machine learning algorithm that we selected in the analytic approach stage, which is decision trees. We can make an observation that:

1.We have categorise the mail according to our need.

2.We have mail as parent key.

3.we have Read and Unread as child key.

4.Unread mail are furthure divided with name of different ids from which mail are coming

5 We can remove if any unread mail is seen in READ category as part of Data Cleaning.

MODELLING and EVALUATION

Here we will train our data to fit into the model .We can check if we are getting the most accurate result using visualization technique as well. We can automate the unsubscription of the particular unused ids as well.Hence we can keep our mailbox free from irrelevant mails and messages.

DATA EVALUATION

Here we can check if the final result is as per we needed or not. If its not most accurate we can check the data again.