# Accenture Data Analytics & Visualization

Presentation summary

* Hello Everyone , My name is Trideep Roy and I’m currently working as a Data Analyst intern for this virtual Data Analytics and Visualization program organized by Accenture North America.
* Before getting into the details , I would like to give a brief introduction about the Analytics team consisting of **Andrew Fleming** as the Chief Technical Architect, **Marcus Rompton** as the Senior Principle and myself as the Data Analyst.
* In today’s presentation I’m honored to present insights tailored to the unique challenges and opportunities faced by Social Buzz in the dynamic landscape of social media and content creation.
* So let us explore Social Buzz, a trailblazer in social media and content creation since 2010. Founded in 2010 in San Francisco by two former engineers, Social Buzz prioritizes content, keeping users anonymous and emphasizing reactions. Over the past 5 years, Social Buzz has reached over 500 million active users monthly, leading to rapid growth and a need for effective scaling.
* But with the rapid increase in their users and with over 100,000 daily content pieces, managing unstructured data has become a significant challenge, necessitating external expertise.
* To address Social Buzz's needs, our 3-month initial project will firstly include an audit of their big data practice, secondly Recommendations for a successful IPO and lastly Analysis of top 5 content categories.
* Moving on from the brief that I just explained , we have used datasets namely Content(**ID,User ID,Type,Category,URL**) , Reactions(**Content ID,User ID,Type,DateTime**) and Reaction Type(**Type,Sentiment,Score**).
* While handling all of these datasets simultaneously ,we came across some difficulties like understanding user engagement and popularity as over 10,000 posts are made per day. Next, it was problematic to capitalize on pieces of content as approximately 36.5 million of them are uploaded every single year.
* Now our whole procedure of Data analysis can be divided into 3 steps.The very first one being Data cleaning,where we remove the missing values,change the data types and drop irrelevant columns.Next we merge the datasets using Reaction as the base table then joining Content and Reaction Type. Lastly we analyze use some visualization to draw conclusions about the challenges that we faced.
* From the visualizations we can see that the observed proximity in values between "Healthy Eating" and "Food" suggests that these categories are often interconnected in the minds of individuals. This correlation could be attributed to the prevalent association between food choices and overall health consciousness. Additionally, it is noteworthy that "Animals" emerges as the most favored category. Possible reasons include the widespread appeal of content related to pets, wildlife, or ethical considerations in animal-related content.
* Thus our 3 main takeaways from this analysis are number 1 JANUARY is the month having highst no. of posts , number 2 there are a total of 16 unique categories and number 3 ANIMAL is the category having highest reactions.
* With this I come to an end of this virtual presentation. I hope that I didn’t overexplain things beyond necessity and you people enjoyed this session as much as I did. Thanks for this opportunity and I hope we’ll cross paths someday or the other. Have a peaceful day bye!