|  |
| --- |
| Hello and welcome, my name is Sanika Khot and today I will be presenting to you the results of the Data Analytics task. |
| Today's agenda will be as follows:  1. We will recap the overall project to give a high level understanding of the business problem we're tackling and the specific requirements.  2. We will dive into the specific problem that we, the Data Analytics team, have been focusing on and will give some background as to why this is such a big problem.  3. After introducing the problem, I will go over the team responsible from our side in tackling this task.  4. I will then go over the high-level process that we followed to complete this task, so that you have complete clarity in how we tackle these kinds of tasks.  5. Finally, I will go over the all important results and I will present them as a series of insights and visualizations from our analysis.  To wrap up, I will summarize and open for any questions. |
| To kick things off let me recap this engagement.  We, Accenture have embarked on a 3 months pilot with Social Buzz to focus on 3 main tasks, aligned with some of the biggest challenges that you're currently facing.  Social Buzz has reached huge scale in recent years to become recognized as a global unicorn company. We are here to help you manage this scale and to guide you in the right direction.  Firstly, we will be doing an audit of your big data practice and sharing best practices and industry expertise. Secondly we will be guiding you through a successful IPO, of which we have deep expertise and knowledge of within our team. And finally, we have conducted an analysis of your data to find insights regarding your top 5 most popular categories of content |
| Focusing on the last point that I mentioned there, this is what the Data Analytics team has been specifically focused on.  Clearly with such grand scale, this comes with a lot of data and with such vast amounts of data comes challenges.  To give a background on how much data you've been creating:  - You told us that your platform receives over 1 lakh posts per day which amounts to 3 crore 65 lakhs posts every year, of which, this is all unstructured data making it very hard to make sense of.  In this day and age, content is king. Just look at some of the biggest platforms in the world, for example YouTube, Facebook and Netflix... they are all content businesses...  But how to capitalize on it when there is so much?  It's not just all about harvesting as much content as possible... The real value is in understanding and crunching this content to gain a deeper understanding of your audience and to therefore provide a more personalized and enjoyable experience.  And this is where our data analytics expertise comes in, with the insights that we've uncovered from this task, we can show you exactly how to take analytics to production at scale. |
| Talking about experience, we have a large data analytics practice at Accenture but we had a team of 3 people primarily focusing on this task. Andrew Fleming is our Chief Technical Architect and his expertise really helped to guide the team to produce high quality analysis.  Marcus Rompton, a senior data expert has worked with the world’s biggest clients on solving their data problems and was heavily involved in the data engineering side of this project.  And finally, myself, Sanika, who was solely responsible for taking leadership guidance and delivering high quality insights from the raw datasets and turning these into business decisions. |
| So, how did we tackle this problem?  Well, we approached it in 5 steps:  1. Data understanding - the key to success on any data project is to understand the data in detail. So we took the time to understand the data model and domain of your business.  2. Data extraction - after understanding your business, we then architected what an ideal dataset should look like for this problem and extracted it from the relevant data sources.  3. After extracting the raw data, we needed to process and model this data into a dataset that can precisely answer the business questions and produce analytics.  4. With our new dataset, we used our analytical expertise to uncover insights from this dataset and to produce visualizations to describe the insights.  5. And finally we used these insights to unlock business decisions and to make recommendations on next steps. |
| From your data we found that you had a total of 16 unique categories of posts across your sample dataset. This includes things such as Food, Animals and Science.  As well as this, there was 1897 posts from just the Animal category alone! People obviously really like animals!  And also the most common month for users to post within was January, since this is such a seasonal month with so many holidays and events, this is interesting to know that people are most active during this month!  But now, onto the main question... which is... what were the top 5 most popular categories of posts? |
| From our analysis you can see that the top 5 most popular categories of posts were animals, science, healthy eating, technology and food in descending order.  Animals had an aggregate popularity score of almost 76 thousand. It is very interesting to see both food and cooking within the top 5, it really shows what people enjoy consuming as content. But also interesting to see science too. Clearly users favor "real-life" content on this platform.  Furthermore technology is an interesting category because it develops each and every day. This presents a huge opportunity for you to differentiate your platform and to run specific content or events linked to this new updates. |
| Additionally, you can see from this chart the % split of popularity between the top 5 categories. There is not much difference between each of them, animals only outperform science by 1.1% within the top 5.  However, the difference between the 4th most popular, technology, and the 5th most popular, food, is much smaller at 0.6%  This tells me that the categories sorted by popularity is weighted towards categories at the top. This means that it exhibits a "greedy" effect, the most popular categories get more popular whilst as you drop down the popularity rankings, you may see that they fall away drastically. |
| So, to summarize:  We tackled this task and found the top 5 most popular categories as asked, but we also went one step further.  - We found that food and culture are the two most popular categories, suggesting that users like "real-life" content  - We also found that soccer was the third most popular, perhaps due to the tournament coming up. This presents a massive opportunity for Social Buzz to ride on this global event, as all eyes will be on it as well as the players.  - As much as this analysis was insightful, we are ready to take it to the next stage and we have the expertise within Accenture to help you realize these kinds of insights in production across your organization and in real time. We would love to help you with this. |
| Focusing on the last point that I mentioned there, this is what the Data Analytics team has been specifically focused on.  Clearly with such grand scale, this comes with a lot of data and with such vast amounts of data comes challenges.  To give a background on how much data you've been creating:  - You told us that your platform receives over 1 lakh posts per day which amounts to 3 crore 65 lakhs posts every year, of which, this is all unstructured data making it very hard to make sense of.  In this day and age, content is king. Just look at some of the biggest platforms in the world, for example YouTube, Facebook and Netflix... they are all content businesses...  But how to capitalize on it when there is so much?  It's not just all about harvesting as much content as possible... The real value is in understanding and crunching this content to gain a deeper understanding of your audience and to therefore provide a more personalized and enjoyable experience.  And this is where out data analytics expertise comes in, with the insights that we've uncovered from this task, we can show you exactly how to take analytics to production at scale. |