Good to you sir. Welcome to our presentation for our COMP301 Project.

I am Talha Vawda, and I will be presenting on behalf of our group.

My fellow group members are…

TOPIC

**Covid19-Digital-Footprint**

Quantifying and classifying the Digital (Social Media) Footprint of South African Telecommunications companies before and during the imposed lockdown as a result of the Covid-19 pandemic,

And correlating it with the transition to remote/online learning in SA

Our social media platform of choice was Twitter. We chose Twitter as it allows for direct interaction between users and companies. (users can mention a company in their own tweet, instead of (in other SMs) only replying on a company’s post) And due to the relatively short length limit of tweets in general (this is one of the unique offerings of Twitter), it makes it easier for us to be able to do sentiment analysis.

**[Companies]**

We chose 4 telecommunications companies to work with: 2 of them, Telkom & Afrihost are primarily ISPs, meaning they responsible for providing Internet access to homes (as WiFi) etc. via broadband/fibre/ even 4G at home.

The other 2 are primarily mobile operators, MTN and Rain. They’re responsible for providing mobile data to customers, giving them access to the Internet on the go, and also for people who don’t have a dedicated WiFi set up at home

**[Introduction] Why we have chosen this topic**

As university students, the Covid-19 pandemic and the lockdown imposed by our government has hit us hard. We’ve had to adapt to online and remote learning, whilst being stuck at home, and our studies required that we have stable internet connectivity. It has been a challenging period for us, and we’re interested in seeing how fellow students and workers who have been WFH have been impacted.

Covid-19 played its hand in forcing telecoms companies to scale and improve their infrastructure to meet the evolving needs of society – specifically with most employees now being made to Work From Home  
So we want to see whether the telecoms companies managed to successfully adapt and upscale their infrastructure to meet the internet requirements of the SA citizens, and how well they’ve done so.

Correlations:

+ online shopping [out]

+ banking [out] (in terms of going digital; staff retrenchment, start ups:bank zero, tyme bank)

+MOOC [out]

* Decided not to focus on it as MOOCS were already popular before covid, there’s not so much of a presence in SA, and from looking at the tweets of MOOCs, there wasn’t much interaction and engagement with users. But we decided to keep the focus in the same sphere, namely education, and focus make the correlation between telecoms and online learning, particularly how SA university students had to deal with the transition to online learning, and whether they’ve been able to manage/cope or they struggled

+ online learning

* We included online learning as part of our data acquisition, however due to time constraints, we unfortunately hadn’t had the opportunity to analyse the online learning aspect with regards to the telecoms companies

QUERIES

*Queries can inspire questions -> from looking at the data, did we find what we thought we'd find (from our speculations) or the sentiment is not intuitive or the opposite (anomalies)?*

Queries that our Web app will generate

* Noted that queries are hard-coded/static – pre-coded

Need to distinguish before/after comparison date

Lockdown Start Date: 27 March 2020

We decided to use the lockdown date as the comparison/transition (before/after) date (instead of when covid was discovered/announced) as this is when everyone had to start working and staying at home, so this will give us a better idea to how well the companies adapted.

End date: 30 June 2020 (3 months of lockdown)

* **Number of Tweets** of each company (in total from start to end date) and also before the lockdown and during/after the lockdown
  + We looked at almost 6 months before lockdown (from 1 Oct 2019 till 26 March 2020)
  + We looked at 3 months during lockdown (27 March to 30 June)
  + Analysis
    - Which companies tweeted the most
    - This query can tell us whether the telecoms companies were more active on Twitter during lockdown as compared to before lockdown
      * Note the time period of before is 2x the time period of during
    - .. Whether theres more interactions with users (being more responsive, connecting with customers, attending to their questions/queries/complaints via Tweeting) and whether theve had more campaigns
* **Most popular keywords and hashtags**
  + Also in terms of the popular keywords/hashtags amongst users, and amongst companies separately.
  + And see if theres any correlation
* The **number of tweets that each company is mentioned in**, before the lockdown and during/after the lockdown
  + Analysis
    - Less users tweeting -> improvement in company services?
* **Top** Twitter **accounts that these 4 companies (frequently) mention** in their tweets
  + Analysis: why so?
* Most popular reasons that the Twitter users complain about regarding the Telecommunications companies
* **Sentiment Analysis of the companies (General)**
  + Sentiment of twitter users towards each company both before lockdown and after lockdown – from looking at the tweets that the company is mentioned in
  + % positive, negative, neutral
  + Analysis
    - Change in sentiment of users
    - Have users responded well to campaigns
* A specific user’s sentiment (given their handle) towards each of these companies based on their tweets
* Sentiment Analysis on the keywords/hashtags
  + E.g. on “Online Learning”
    - How the users feel about this topic.
* How SA university students have been affected by (their sentiment with respect to) transitioning to online learning/studying and whether they’ve been able to cope
* SA universities [**see slide** Queries 2] and if they offering online courses and stuff before/after Covid and the sentiment of students with respect to transitioning to online learning

WHAT WE DID (INCL. ARCHITECTURE)

**Data Acquisition**

* Used Python with Pycharm
* Tweet scraping/crawling from Twitter – Twint API
  + We stored the data in 2 csv files
    - Data.csv – the tweets from the 4 companies during the timeperiod
    - Keywords.csv – the tweets that contained the keywords and hashtags that we specified
      * Some of them are: …
* We set up a script that extracted tweets from our 4 telecoms companies during the entire time period: Telkom, Afrihost, MTN and Rain. This is the data.csv
* We defined a set of keywords and hashtags that we are interested in, to filter the tweets that users tweetes. These keywords deal with online learning, working from home and the general technology struggles that was experienced during the lockdown.
* we were able to get tweets from individual user accounts that mentioned either of these companies in their tweets along with one of the keywords specified. This is the keywords.csv file. Note that a user replying to a tweet from a company is also included as a mention
* CleanData
* API Options
  + *Regarding our research, we checked out the ffg Twitter APIs, official Twitter API, Textblob, Twint and and Tweepy. Textblob seemed more aimed at sentiment analysis of the tweets instead of actually getting the tweets so we didnt want to use that.*
  + *Tweepy but it needs the Developer License thing from Twitter (which we tried acquiring but had to verify ourselves using a cell number; didn’t do it for privacy reasons) so it seems like Twint may be our best bet, as we just had to import the API into our code*
  + [limitations] Number of actions we can perform is limited with the Twitter API

**Database**

* We chose Neo4j as a our NoSQL Graph Database
* Learned and used Cypher language for queries to create the database, and also to formulate queries to the database
* 6 Queries to create the database
* Nodes in the database:
  + Company
  + User
  + Tweet
  + Hashtag
* Every member had to create their own local database. When we run the web interface, it will connect to our own local database (This was because of some technical difficulties we were having with setting up the DB as a server, which I will mention in due course). Each person having their own local db is fine as the data/model is read-only / static. We will not be modifying it.

**Web Interface**

* The project of our web app is called WebInterface
* It is a dashboard that lets you navigate to the queries you want to view (each of these queries activates a Controller)
* It is not ideal to split up the code, but since we working as a group, we made each query have its own controller, so that we can work on them in parallel
* Jetbrains IntelliJ IDEA IDE
* Java, Javascript, HTML, CSS
* We chose to use Spring Framework as the Web Framework that we’d use to build our web application/interface
* Cypher queries are in QueryNexus

**Design Patterns**

*See shaelins notes*

**MVC Design Pattern**

* Model is the database (and APIs used to contain it - Neo4j has it)
* View is the resulting (generated) HTML
* Controller is either a servlet or JSP that performs the query and then constructs the View (actual webpage)
  + Takes in input from the UI to execute query, get the result, and send the data back
  + Can have 1 Controller for all the queries or a Controller for each query. We have to decide this granularity

**Collaboration**

GitHub

TECHNICAL DIFFICULTIES

**Database**

* Neo4j
* Cloud – MS Azure
  + GitHub Student Pack

**Web Interface**

* Exception handling – error page

WEB APP DEMO

RESULTS/ANALYSIS/CONCLUSIONS

* What type of **Analytics** we did
  + Tabulated data
  + Line graphs
  + Node graphs, -> showing companies, hashtags
  + Sentiment Analysis

*From looking at few entries in the csv files, the following trends were seen: (Formulate queries based on them and add them to the Queries list)*

* *Users on twitter are complaining/ranting about the service provided by the ISPs/ mobile networks*
  + *Network connection – no connectivity, v low speed*
  + *Service delivery*
  + *Poor customer services*
  + *Data disappearing*
* *Asking about products/prices*
* *Students complaining about network coverage*
* *A lot of the users’ issues are with Telkom*

*Consider the bias that users are much more likely to express a negative sentiment than a positive sentiment (as they'll only tweet when they have an issue)*

*'No news is good news'*

*If there is a drop in complaints then you can take it as a positive outcome*

*More complaints could also be seen as customers opening up to them and coming forward with their complaints/dissatisfactions*

NOTES (ISP vs Mobile) for Conclusion

I guess. But I was thinking that from start of lockdown the mobile data usage woudlve remained somewhat constant but the usage of ADSL/fibre wouldve increased cos everyone is at home

Hmm yeah that should be the case, but it depends on the level of development of internet infrastructure. in poorer countries with underdeveloped infrastructure people tend to rely on mobile internet.

True. My thought process was that those relying solely on mobile data for Internet access are probably not white collar workers and whose jobs dont rely on them using the internet. But as you saying, since they were more at home during lockdown (and not working) theyd be using the Internet (via their mobile data) more

We will also have to take into consideration when doing our analysis that during the lockdown period, tertiary students (around couple 100000) wouldve migrated their internet access from through their uni's ISP to mobile providers (by using the mobile data provided). We can see if the non-usage of the institutions Internet made any tangible negative impact on the ISP internet usage

And to counter my own point here, some of those who would normally rely on mobile data might have needed to cut off their internet to save money when their jobs got put on hold