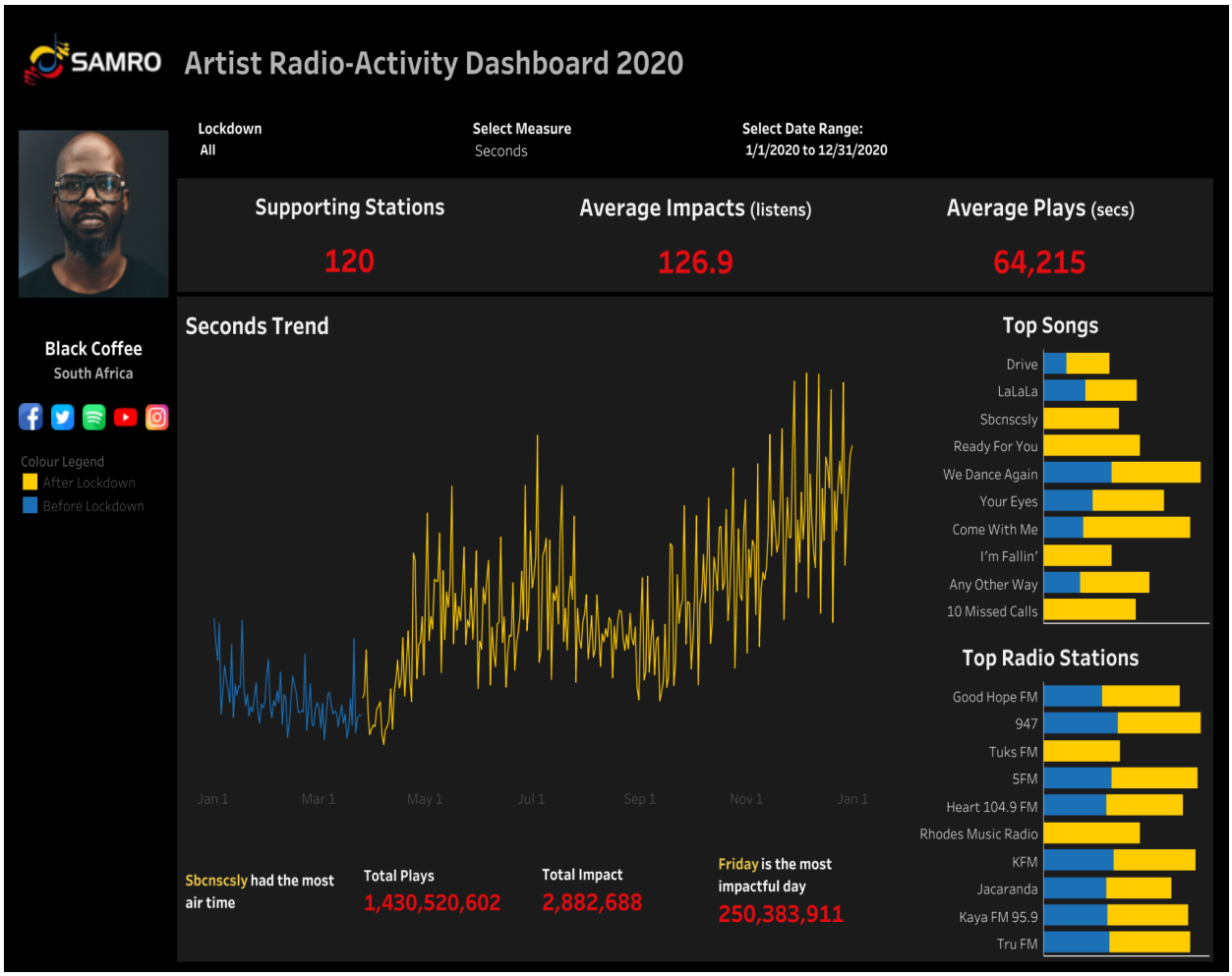
CETM25

Artist Radio-activity Dashboard Report

Mark Stent - 4 January 2022

Wordcount: 2510



Introduction

This report covers the process and management of the SAMRO (South African Music Rights Organisation) Radio-activity dashboard I have created as a proof-of-concept design for this module. I was permitted to use a one-year time frame, as it would provide better insight into the data and the problem. Please refer to my first assignment in Appendix 5 for this project's initial goals and objectives. The dashboard is available here:

https://public.tableau.com/views/BlackCoffeeAnalysis/ArtistRadio-ActivityDashboard2020?:language=en-US&publish=yes&:display_count=n&:origin=viz_share_link

All data used in this project has been collected and used according to the GDPR (2016/679) and the Data Protection Act (1998).

Section 1

Selection Of The Visualisation Software And Hardware

From a software perspective, the project breaks down into two layers: Exploratory data analysis (EDA) and visualisation/dashboarding.

Python is an open-source language that is excellent for analysis work. It is widely accepted, extremely portable, and offers various libraries to perform any required task. Libraries are reusable chunks of code that assist in speeding up or simplifying problem-solving. These libraries are the main reasoning for using Python in this project. The specific libraries used are listed below:

- Numpy offers a comprehensive suite of mathematical functions and tools used in this project for statistical analysis and other analysis functions.
- Pandas is a data manipulation and data analysis tool. This project uses Pandas to collate, manipulate, and manipulate the dataset for analysis.

- Scipy is a library offering fundamental algorithms for scientific computing. This project uses Scipy for statistical analysis.
- Statsmodels offers classes and functions that cover many statistical models. For example, this project uses Statsmodels for time series analysis.
- Outlier analysis was done with Facebook Prophet.

Jupyter is an integrated development environment (IDE) used to compile code in an 'all in one place' manner, allowing equations, images, plots, and code to be hosted in a single interface; this allows for faster analysis and exploration of the data in a data science project. These qualities make Jupyter notebooks ideal for this project.

Tableau handled the visualisation/dashboarding component. Tableau is widely used for business intelligence and is known for creating powerful visuals backed with excellent support for big data. It is also relatively easy to use and understand. In addition, Tableau has a public version that allows the hosting of smaller projects on a shared server with all of the enterprise features enabled; this made it ideal for use in this project.

Product Development / Software Engineering Methodology

Agile is a mindset that enables users to adapt to an ever-changing environment. Scrum is a lightweight framework that helps people, teams, and organisations generate value through adaptive solutions for complex problems in an agile environment (scrum.com, Nov 2020). Since this project is exploratory, the agile approach fits perfectly.

The scrum methodology calls for short sprints of time-boxed work to complete a defined amount of work and an organised process to manage time and project outputs. Accordingly, I followed the below agile stages:

- Stage one - Requirements gathering and product backlog creation
- Stage two - Exploratory data analysis

- Stage three - Tool development and creation

Since this was a small project with few stakeholders, all three stages were assigned to one sprint of 4 weeks. The project status underwent daily review, updating current progress iteratively in the backlog file (Appendix A). The sprint review meeting would decide the next steps for the project.

Radiomonitor.com provided the original data in a .CSV file. This data underwent the following pre-processing to clean the data for analysis and output into Tableau:

- Pandas was used to create a data frame.
- Date column converted to a date/time datatype.
- Leading and trailing whitespace removed from around the data
- Remove all channels where impacts = 0 as they are assumed not to have official RAMS figures.
- Impute null values with 0's
- Convert 'impacts' and 'seconds' columns to numerical data types
- Get rid of rows that don't make sense, i.e. where there are no seconds of play but have > 0 impacts (this could be due to data entry errors)
- Changing of column names to more readable names
- Statistical analysis of data to check for any statistically significant events and changes over the period includes tests for normal distribution, Bartlett's test for differences in variance and t-test for difference in means (Appendix 3).

System Testing

Testing of the dataset broke into two phases:

- Phase 1 - Testing and exploration of the raw data set for insights (see Appendix 3)
- Phase 2 - Testing the final visualisation

The goal of phase 1 was to do a deeper dive into the data, run some initial statistical tests to check the shape and spread of the data, and do a time series analysis to check for any seasonality or outliers.

The first test checked if the 'seconds' and 'impacts' columns fitted a normal distribution. Using a combination of analysis histograms, QQ plots, kurtosis, and skewness metrics, we can assume that 'seconds' follows a normal distribution, while 'impacts' does not.

Bartlett's (Snedecor and Cochran, 1983) test for significant change in variance before and after the lockdown was tested and found to be significant, implying that Black Coffee's airtime increased after the lockdown period. However, more analysis is needed to decide if the lockdown was the reason for this increase.

A t-test (Box, Joan. Fisher, 1987) for change in means after the lockdown date was performed on the data and was found to be significant. This test implies that mean seconds played increased after the lockdown date; this could be due to more people listening while at home during the lockdown. More investigation is required.

Time series analysis showed an overall upward trend in airtime. A definite seasonality exists on Saturdays; this makes sense as weekend radio shows have more freedom in playing tracks for extended periods, with less talking.

A Facebook Prophet model using 99 percent confidence intervals of predicted means was fitted to detect outliers. Only seven outliers occurred, and all on Fridays.

Phase 2 used the cleaned and processed dataset after testing, decreasing the probability of data-related errors in the visualisation tool. In addition, each feature was tested using every

permutation of interactivity on the visualisation tool for 'breaking' and accuracy. As a result, no visible errors or software errors occurred.

For production versions of the tool, a layer for error checking and 'broken data' needs to be developed.

User Evaluation

This project was conducted using Scrum methodology. In Scrum, the sprint review is employed to review the outcomes of a sprint and determine future adaptations or plans. Inviting the product users to this review is an excellent way of validating the project features and results.

The best way to get honest feedback from users is to hand them the tool and let them use it. Then, after they have had time to 'test drive' the product, feedback is collected.

Feedback is documented by asking questions to the users that align with the original project goals, for example:

- Do you understand what the dashboard is depicting?
- Is the layout intuitive?
- Is there anything you don't understand or doesn't make sense to you?
- Do you find the experience pleasant/unpleasant? Why/why not?
- Do any new ideas come to mind that could be added to the tool to make your job easier?
- What insights did you get from using the tool?

The vital part of collecting this kind of information is to make sure that the user feels safe expressing his opinion, setting clear goals for the meeting, and keeping the interactions light-hearted and non-judgemental; this will encourage honest and objective feedback.

Matching the results of this review session against the initial project goals would dictate the success of the first sprint. Any changes or new requirements would be updated in the project backlog and included in the plan for the second sprint.

Since the primary goal of this project was exploratory, the results of the first sprint review would decide the direction of the project going forward.

Section 2

Time Management

I am currently working full time as a data analyst and doing my master's degree part-time; I also have three daughters and a family. With this in mind, time management has been a massive influence on the quality of my work output. However, with the decision to complete this assignment with an agile mindset and using Scrum processes, achieving timeous goals was made much more straightforward. The Pomodoro technique was another tool I used to help me meet my goals.

The project backlogs (Appendix 2) main aim is to document project outcomes, assign priorities, and estimate the effort required for each user story. I used this document to plan the time I spent on each story and tried to keep it within the target hours as much as possible. In addition, because each story was assigned a priority, I was able to focus on the most critical tasks first. If any tasks overshot the estimation, I re-adjusted the estimates based on what was needed.

The Pomodoro technique was developed in the 1980s by Francesco Cirillo to encourage users to work with the time they have rather than against it.

The technique aims to achieve total concentration with minimal distraction for short periods; these periods are affectionately named 'pomodoros'. I blocked my time into 25-minute 'pomodoros' and 5-minute breaks. During each 25-minute block, I allowed no distractions from social media, phone calls or anything else. I kept track of these sessions and aligned them with the estimated hours in the project backlog.

Overall, this process worked very well, and I was able to keep within close distance of the original backlog estimations. I put this success down to planning and using the Scrum methodology to achieve the project's goals.

If I had to redo this project, I would spend more time planning and framing the questions before jumping straight into the analysis.

Risk Assessment

The project's initial goal was exploratory; therefore, the overall risk is very low, but this would change if the proof-of-concept goes into production. After discussions with the stakeholders in the first sprint planning meeting, highlighted the following risks:

- Accuracy of data - inaccurate data could lead to wrong decisions. Radiomonitor provided the data and has assured that the data is around 95% accurate. There are some minor anomalies where the algorithm detects plays under one second and tracks them as impacts. The data pre-processing cleaned out these anomalies.
- Choice of the artist - Black Coffee is a well-respected artist, and his radio airtime support may not be representative of the South African radio market. In the next sprint, other artists will be tested and added to the dashboard that are more representative of the typical South African artist population.
- Radio station outstanding payments - not all radio stations are up to date with paying royalties, and the figures used in the tool may not translate to accurate royalty payments. If required, the team will undertake a more profound analysis in the next sprint to assess the impact of non-paying radio stations.
- Complexity and cost of the next phase of the project - Depending on how the requirements evolve for the final product, complexity and expenses for the technology required to deploy the visualisation tool may be too expensive.

Quality Control

Using Scrum as a project management methodology requires constant evaluation and re-evaluation in an iterative process at each project stage.

Building plans for quality assurance to the customer and quality control for the project is critical for meeting project deliverables. Using various metrics or questions at each iteration will keep the quality goals of the project in check. Questions include:

- Is the visualisation tool meeting the initial project goals?
- Is the visualisation tool doing what it is supposed to do?
- Is it responding to the questions it is supposed to answer?
- Is the customer satisfied with the output?

In each feedback meeting, these questions are repeated to ensure that the project is going in the correct direction. When a quality problem is identified, the process is reactionary, and solutions to the problems are suggested—for example, modifying the path rather than completely recreating some sections by catching deliverables that don't satisfy the agreed-upon standards is more efficient.

In short, for this project, quality control standards were measured by how closely the visualisation tool met the project goals.

Customer Relationship Management

In a Scrum environment, customer management is managed through the project backlog (Appendix 2), which contains a list of outcomes for the project. In addition, the project backlog provides a system for cleaning up customer service concerns: defects and problems are managed within the sprint period. Issues are measured, reviewed and improved.

Each iteration of the visualisation tool would involve the SAMRO and stakeholder teams, addressing potential issues and questions. Any changes would be added to the project backlog to manage current or future sprints.

Product Market Strategy

The first sprint review discussed the results of the initial exploratory analysis. The project goals were as follows:

Functional requirements:

- Visualise changes in spin data and radio impact data, enabling users to view changes over time.
- The tool should allow filtering individual radio stations or subsets of stations.
- Create a view to visualise the distribution of spins (measured in 'seconds played') across radio stations for individual titles to see the changes in radio support of titles.
- A view of summary spins per weekday to see how listenership is affected by the day of the week.
- Annotate visualisations with essential dates such as lockdown start date.

Non-Functional requirements:

- Views should be easy to use and uncomplicated to understand.
- The visualisation should be secure and reliable for the reporting to be accurate.
- The tool should cover one year starting on 1 January 2020. The first South African lockdown began on 26 March 2021.
- The date range must be interactive.

The visualisation tool met each of these requirements. But, unfortunately, the visualisation tool did not give a definitive answer to the question about lockdown radio support for Black Coffee. Therefore, deeper analysis (possibly a change-point analysis) and more data are needed to answer this question accurately. While Black Coffee's radio support has increased post lockdown, other factors may have contributed to this, such as releasing a new album, media attention and record label plugging into the radio stations.

Despite the visualisation tool not answering the initial question, the SAMRO team have seen other applications for the visualisation tool that could help them with more comprehensive business intelligence. Some of the applications and benefits are:

- Trends regarding radio station spins and impacts
- Trends regarding artist
- Trends in music genre support
- The tool would be helpful for funding and advertising discussions.
- The information could be used to obtain sponsorship for artists from marketers and advertisers.
- The dashboard could assist artists in future music and career decisions.
- The dashboard is helpful for discussions regarding royalty charges for radio stations.
- Radio stations do not always report accurate spin numbers. The visualisation tool could assist in reconciling actual versus reported figures.

The SAMRO team has decided to allow me to develop this project further, and I hope to put the final dashboard together for them by the end of the first quarter of 2022.

Appendix

1 - References

Box, Joan Fisher. “Guinness, Gosset, Fisher, and Small Samples.” Statistical Science, vol. 2, no. 1, 1 Feb. 1987, 10.1214/ss/1177013437. Accessed 5 Apr. 2021.

Jones, Douglas H. “Statistical Methods, 8th Edition George W. Snedecor and William G. Cochran Ames: Iowa State University Press, 1989. Xix + 491 Pp.” Journal of Educational Statistics, vol. 19, no. 3, Sept. 1994, pp. 304–307, 10.3102/10769986019003304. Accessed 2 Apr. 2020.

“NumPy.” Numpy.org, 2021, numpy.org/. Accessed 28 Dec. 2021.

“Pandas - Python Data Analysis Library.” Pydata.org, 2021, pandas.pydata.org/. Accessed 28 Dec. 2021.

“SciPy.” Scipy.org, 2021, scipy.org/. Accessed 28 Dec. 2021.

“The Pomodoro Technique® - Proudly Developed by Francesco Cirillo | Cirillo Consulting GmbH.” Francescocirillo.com, 2020, francescocirillo.com/pages/pomodoro-technique. Accessed 29 Dec. 2021.

“We’re Changing the Way You Think about Data.” Tableau, 2019, www.tableau.com/trial/tableau-software?utm_campaign_id=2017049&utm_campaign=Prospecting-CORE-ALL-ALL-ALL-ALL&utm_medium=Paid+Search&utm_source=Google+Search&utm_language=EN&utm_country=SA&kw=tableau&adgroup=CTX-Brand-Core-EN-E&adused=504564000692&matchtype=e&placement=&gclid=Cj0KCQiA5aWOBhDMARIsAIXLlkd0j5X_GXBAI2fygP9VhGeHzmgDQFZNTeYpgYtHgwip3SCmzmxi_YcaAqUyEALw_wcB&gclsrc=aw.ds. Accessed 28 Dec. 2021.

2 - Project Backlog

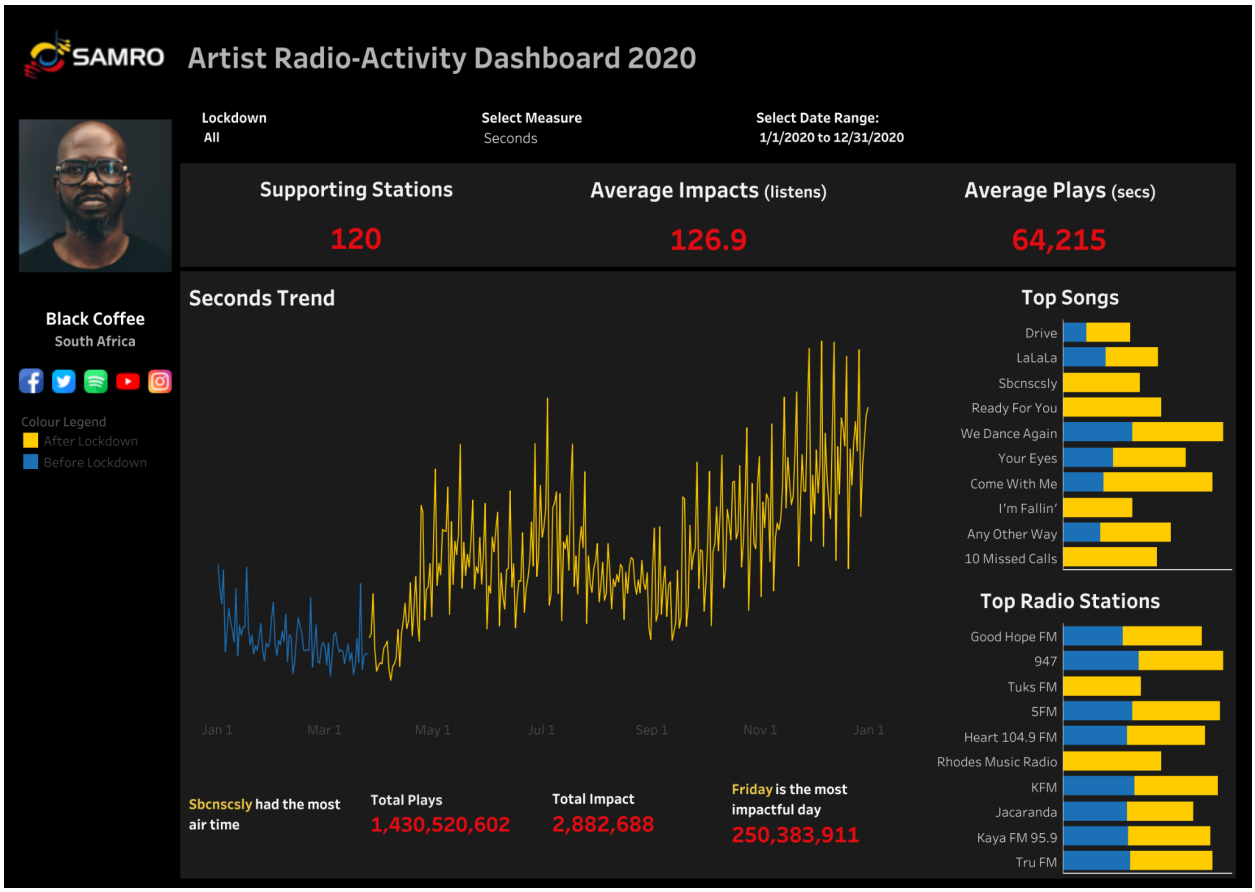
Filetype: Numbers
Filename: SAMRO-tool-backlog.pdf

3 - Pre-processing Python Jupyter Notebook

Filetype: PDF

Filename: BC - Datacleaning.PDF

4 - Dashboard screenshot



5 - Assignment 1

Filetype: PDF

Filename: Masters visualisation report 1.pdf

6 - Glossary

RAMS - Radio Audience Measurement - A metric used to measure radio impacts

SAMRO - South African Music Rights Organisation

SPIN - A single play of a music track

SECONDS PLAYED - The total number of seconds of play of a track for a specific spin.

IMPACTS - A metric created by The Broadcast Research Council of South Africa (BRC), a non-profit organisation. 'Impacts' tracks the influence of the radio station by listenership and demographic using Radio Audience Measurement (RAMS) figures.