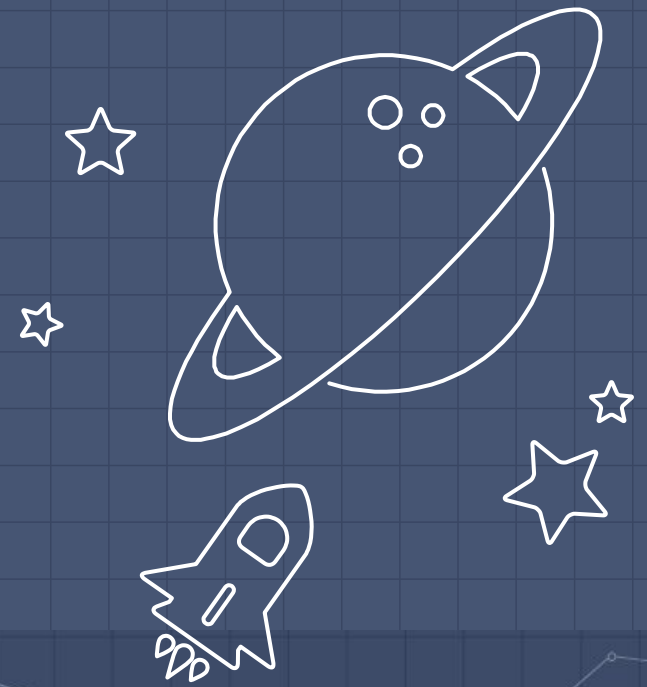
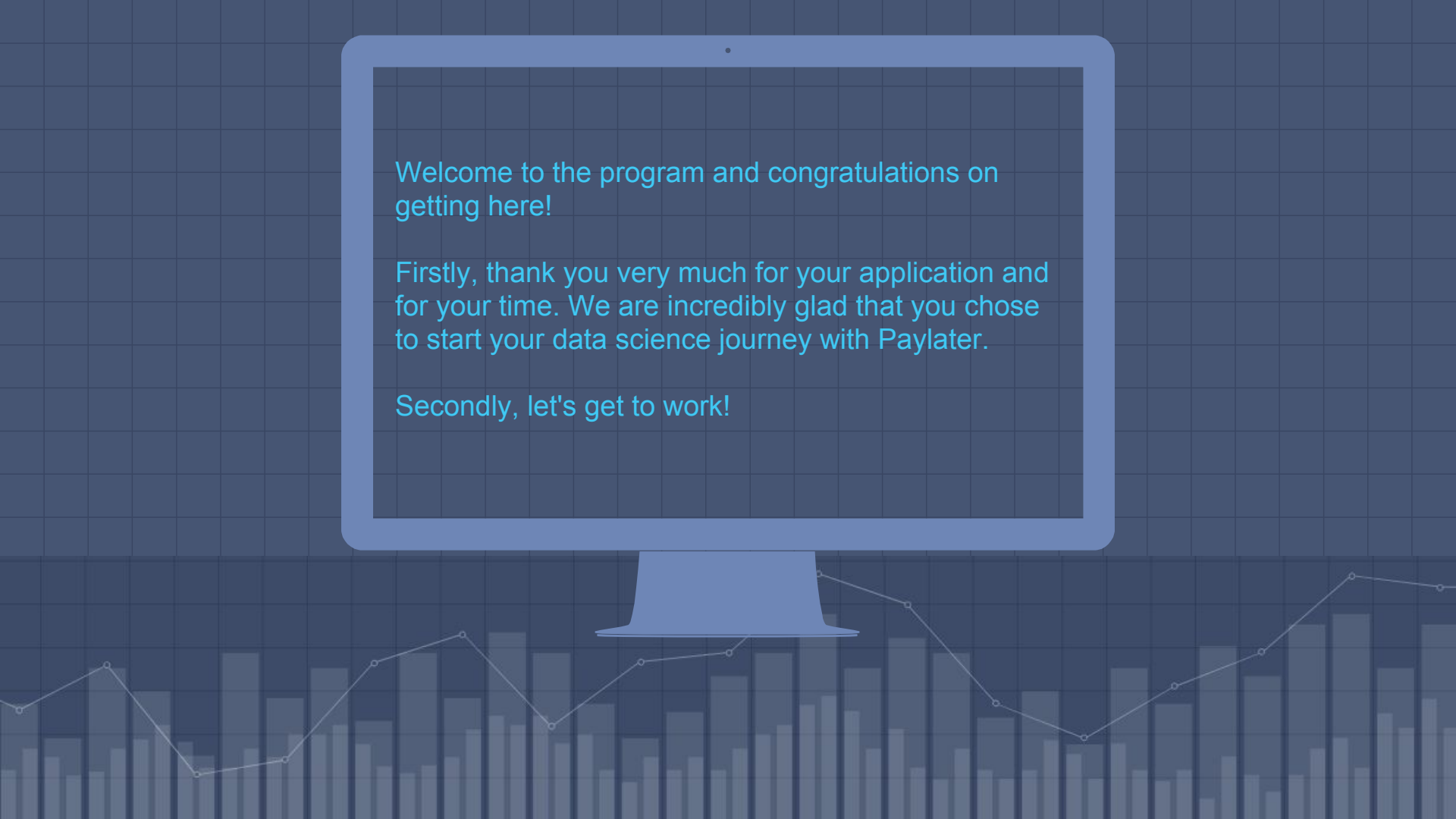
arise:

Data Science Fundamentals

Week 1




Welcome to the program and congratulations on getting here!
Firstly, thank you very much for your application and for your time. We are incredibly glad that you chose to start your data science journey with Paylater.
Secondly, let's get to work!

A Little From Our Co-Founder: Ngozi Dozie

"Thank you for taking the time to participate in the Arise Data Science Internship. We live in exciting times where with the advent of technology, cheap bandwidth and access to free, online courses, anyone is able to transform themselves into a data scientist. This is important to ensure that Nigeria and indeed millions of others in Africa don't get left behind in a world that is getting more digitized with each passing day.

Machine Learning and Artificial Intelligence are here to stay - there is no doubt about that. What is uncertain is whether Africa will be left behind and this is where you have a big part to play. For Paylater, people with strong data science skills are a critical resource - just like water. Arise is one way to help train the next generation of data scientists and also support the fantastic work done by DataScienceNigeria in training the next generation of knowledge workers.

Our hope is that this internship will help inspire others to become data scientists and also form a pathway for those looking to work with Paylater."

Our Data Science Team

Pascal Bernard: Head of Data Science

"I completed a data science graduate certificate at Harvard University and have worked as data scientist for the last 4 years.

Working at Paylater and leading the data science team has been a unique challenge for me. We have many talented people working in the company and have valuable experience in transcending what data science can bring to Credit Scoring, Risk Analytics, Portfolio Optimization and Management, and Fraud detection and Prevention."

Jacobo Varela: Data Scientist

"I am a data scientist with a background in chemical engineering.

Working for Paylater has been a great challenge so far, working with very talented people on interesting projects like new scoring models to understand our customers, and designing the decision engine with which Paylater offers loans.

It is very exciting to see how data science is being used to drive change within the company, and the team is really motivated to use the latest technologies to offer the best service possible."

Devon Stone: Data Scientist

"I am currently the final year of my masters in data science at the University of Cape Town and as a result, I really understand the value in a program like arise.

I have worked for Paylater for a year and have learnt an incredible amount. I love my job and the freedom that comes with it. At Paylater we are trusted to make decisions and encouraged to take on responsibilities.

One of the major reasons I decided to join Paylater is because I believe in what we do and the goals we have set out to achieve."

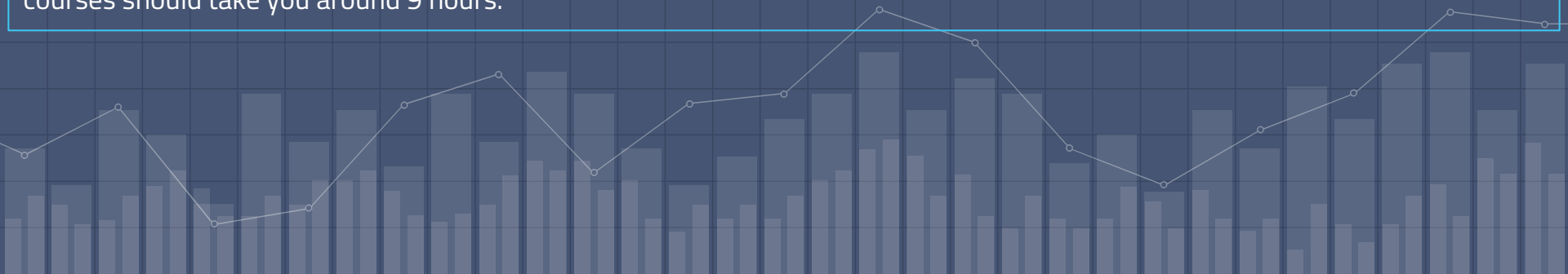
Coding Introductions

Each week of the program you will be given readings and possibly links to external coding sources relating to topic of the week.

To begin this week we would like you to go through the following introductory courses. If you have gone through them before then you can move onto the next section but we do advise revisiting them to sharpen up your skills.

1. Data Camp's Introduction to Python.
2. Data Camp's Introduction to SQL.
3. This Introduction into GIT (this well help you understand how to use Github, which is a key data science skill and great for your online data science presence).

The Data Camp courses are free, if you are asked to pay then you have selected the wrong Data Camp Course. In total, the three courses should take you around 9 hours.



Readings

We ascertained from the application replies that our applicants come from a wide variety of backgrounds. We further gathered that many of our applicants were a little unclear with a few key data science concepts. As a result we would like you to read through some introductory data science readings. These readings should provide you with some fundamental knowledge required for the industry.

1. Statistics reading from Towards Data Science.
2. Supervised vs Unsupervised from Towards Data Science.
3. Classification vs Regression from Medium.
4. Beyond just Accuracy of a classification model from Towards Data Science.
5. Machine vs Deep Learning from KDnuggets.
6. ML in Credit Scoring and Fraud Detection from Forbes.

You should have a better understanding of data science and how it can be applied to the lending and banking sectors.

Data Science Life Cycle

paylater

The data science life cycle can be summed as all the processes that are involved in a typical data science project.

One of the great things about working for Paylater as data scientists is that we get exposed to the entire data science life cycle throughout our time at the company.

1. Define Project Objectives

- Define the problem being tackled.
- Define the features that will be used in the models and the target variable.
- Prioritize modeling criteria. For example does the speed of training the model outweigh the importance of of the model's accuracy.

2. Wrangle and Explore Data

- Locate data.
- Merge data into one table.
- Analyse data.
- Develop any features required for the model.

3. Build the Model

- Select the features that will be used in the model.
- Build the model.
- Validate the model.

4. Interpret and Explain Model

- Try to understand the model and the impact of each feature used on the model.
- Explain this to the team and to other members of the company.

5. Implement and Maintain Model

- Host the model using an API framework.
- Test the API endpoint and what the endpoint returns.
- Monitor and maintain the model.

You might be familiar with some of these processes, but do not worry if you are not! They will all be explained throughout the program.

It is key to note that you might not always be required to run through the entire cycle on a specific project. You might be tasked with wrangling the data required for a model or setting up the API endpoint required to host a model.

Next Week

paylater

This week was an introductory week, as of next week you will be allowed to sit with and manipulate real world data. Next week we will be dealing with section 2 of the data science life cycle.

You will be shown an SQL code example and then tasked with creating your own SQL query that will require you to pull data from multiple tables. You will then have to manipulate and group the data and then display your findings on Google Data Studio (GDS).

GDS is an open source data visualization tool. It would be a good idea to familiarise yourself with the tool before next week.