DIGITAL NATIVE

AI Churn Model Challenge

Purpose

The purpose of this challenge is to separate the strong applicants from the rest.

A strong applicant will:

- Follow instructions,
- Pay attention to details,
- Be able to explain how they have worked logically to solve a problem,
- Use AI in imaginative ways,
- Go above and beyond,
- Understand what they are doing!

You are expected to use Generative AI (GAI) to help you complete this challenge, you will be expected to present at interview how you have used GAI.

Background

This challenge will help you become familiar with the role of a data engineer and the types of things they have to do.

(Hint, it involves solving problems when things don't work as they should!!!)

A Data engineer will typically do the following:

1. **Data Pipeline Development**: Create and manage pipelines that extract, transform, and load (ETL) data from various sources into storage systems like databases or data warehouses.

2. **Data Storage Design**: Design and optimise databases, data lakes, and warehouses to store data in structured and accessible formats.

3. **Data Integration**: Combine data from multiple sources, ensuring it's clean, consistent, and ready for analysis.

DIGITAL NATIVE

Overview.

This activity is designed to take 4-5 hours, but everyone is different, how long you take is not important. Our best applicants return this challenge in 5-7 days. Once completed this activity will be used to screen candidates and forwarded to employers who may use it for a discussion topic at interview.

To complete this activity, you don't need experience of writing code or developing solutions as everything can be searched for, Generative AI is may be used to help, but remember you will be asked to explain how you have used it, and how you validated anything it produced at interview...

We are more interested in **your** ability to follow instructions**, your** attention to detail, **your** ability to use AI tools innovatively and **your** desire to solve the problem than if you can blindly cut and paste…

In this activity you will:

- Task 1: Install Anaconda and Jupyter Notebook, if you don't already have it,
- Task 2: Download a dataset from a public source,
- Task 3: Sign up for a Free Snowflake Account
- Task 4: Open all of the files you've downloaded and combine them into a single data set.
- Task 5: Identify and address any data quality issues, i.e. Missing fields
- Task 6: Create a Cloud Database, and Load the cleansed data into it,
- Task 7: Create a Presentation (i.e. PowerPoint). For each of the Tasks/steps below take a screenshot of the output and reflect on:
  - What you've learnt,
  - Why it is important,
  - and what you might do differently next time.
- Task 8: Upload your challenge.

If you need them here are some useful links (the links are the underlined text, just click on them):

- An introduction to Jupyter
- An introduction to loading data into Snowflake with Python
- An introduction to Prompt Engineering

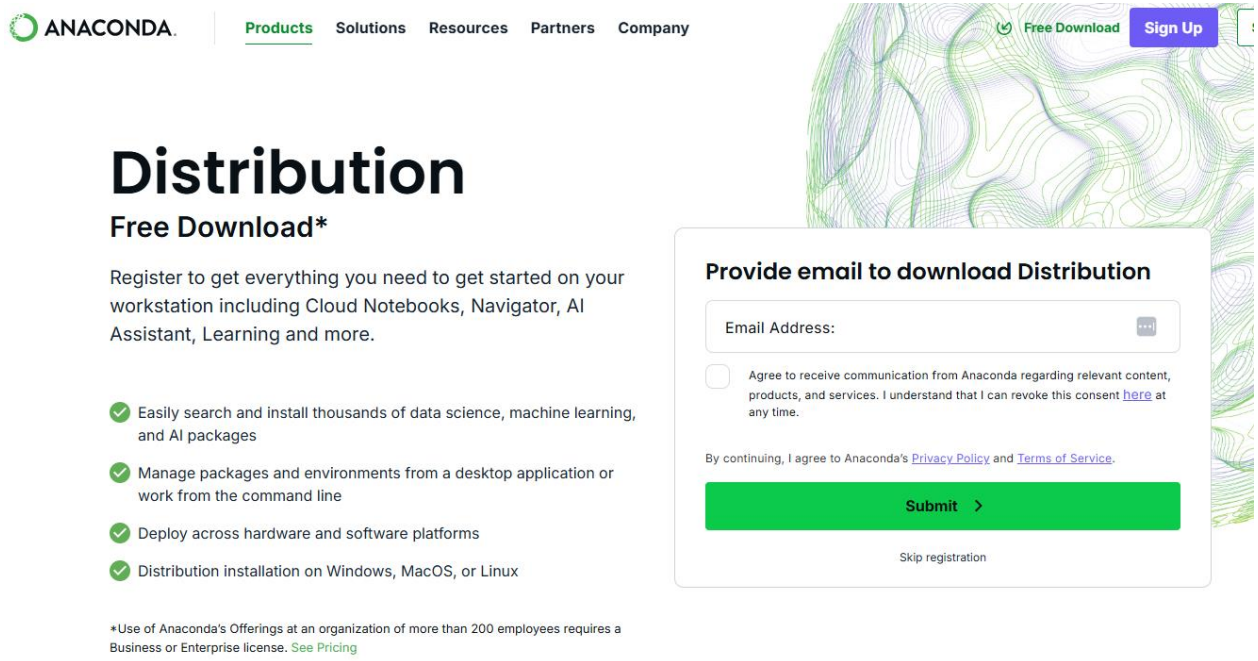
Task 1: Download and Install Anaconda and Jupyter Notebooks

Jupyter Notebook is an easy way to start programming in Python is it widely used in the Data community.

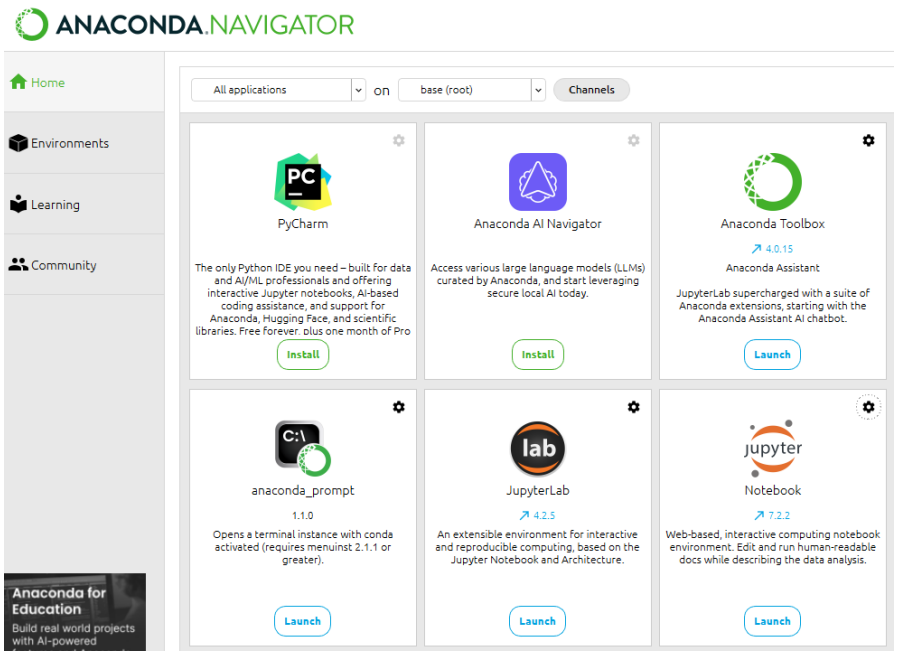
It ships as part of the Anaconda distribution.

Click on this link, enter your email address and press submit to download Anaconda:

Download Anaconda Distribution | Anaconda



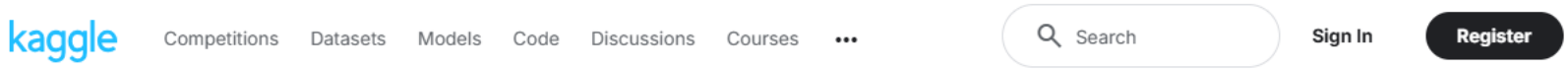
Once the process is complete, you can access Jupyter Notebook from within Anaconda.



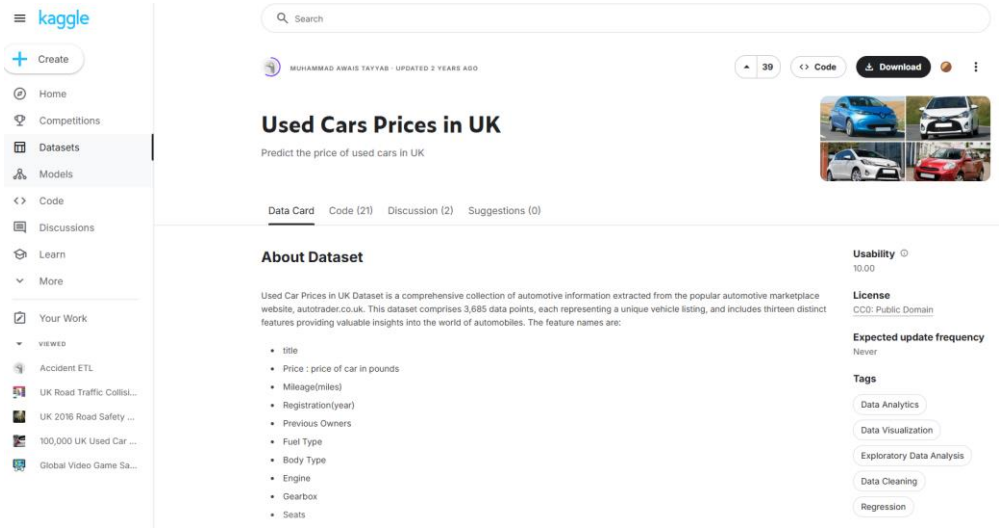
Use AI to learn how to ensure you have the Pandas library installed and include the results in your presentation.

Task 2: Download The Dataset

To get the data for our pipeline, create an account at Kaggle.com,



Click the image below and download the file:



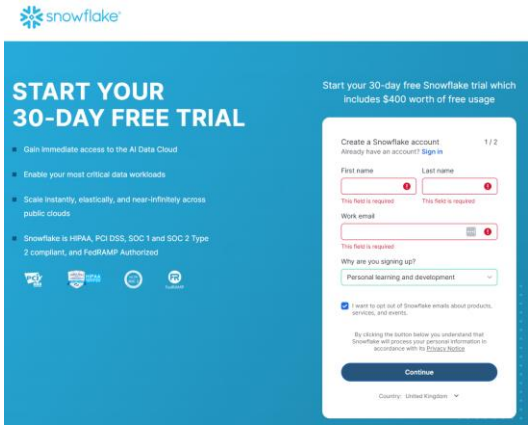
Alternatively follow the link here: Used Cars Prices in UK

Optional Stretch Challenge:

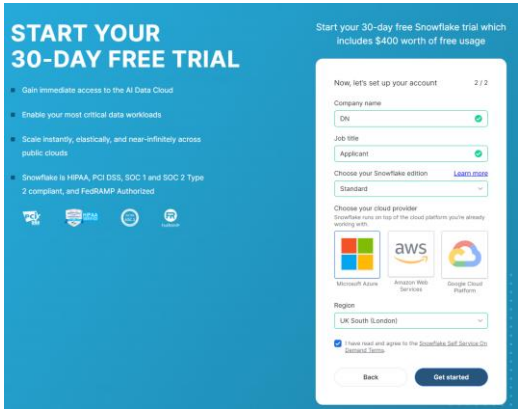
Can you import the dataset directly into Jupyter Notebook from Kaggle?

Task 3: Sign up for a Free Snowflake Account.

Click on the image below, or here to create your own snowflake account

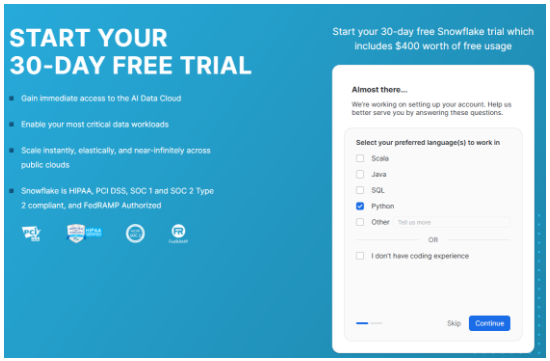


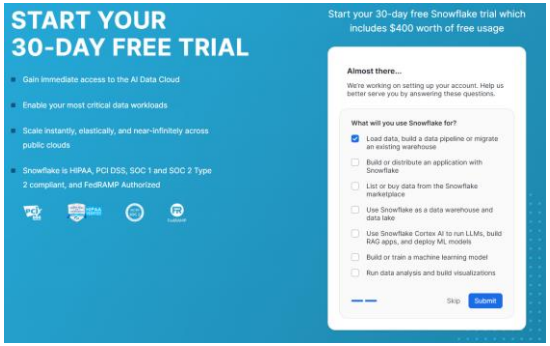
Set you location to the UK.



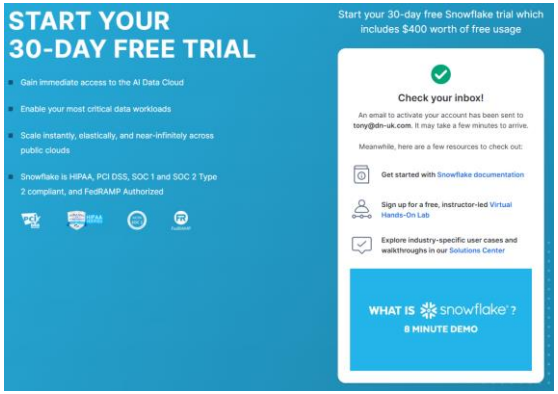
Select Python as your language

DIGITAL NATIVE





Finally follow the link to the documentation,



Task 4: Open and Combine the files you have downloaded

load the data frame into Python using the Pandas library.

Below is an example for how to import a single file, work out how to import and combine all of the files you downloaded and include what you have done in your presentation.

```
import pandas as pd
```

```
df = pd.read_csv("audi.csv")
```

You can check what you have imported with the Head() command.

```
df.head()
```

Optional Stretch challenge:

What other information can you find out about your data set?

Include screenshots in your presentation.

Task 5: Identify and Address any Data Quality Issues

Common Data Quality issues include:

- Missing data,
- Data of the wrong type or format,
- Outliers i.e. values that or much higher or lower than the rest of the values.

Research common data quality issues and how to identify them in Python.

Include in your presentation the Data Quality issues you found.

Research how to remove any remove any records with data quality issues from your combined data set.

Include in your presentation the records you removed from the data set and why.

Task 6: Create a Cloud Database

In this task we will connect to the account you created earlier, create a database and load your combined data into it.

First of all, include the Snowflake Snowpark package into your environment.

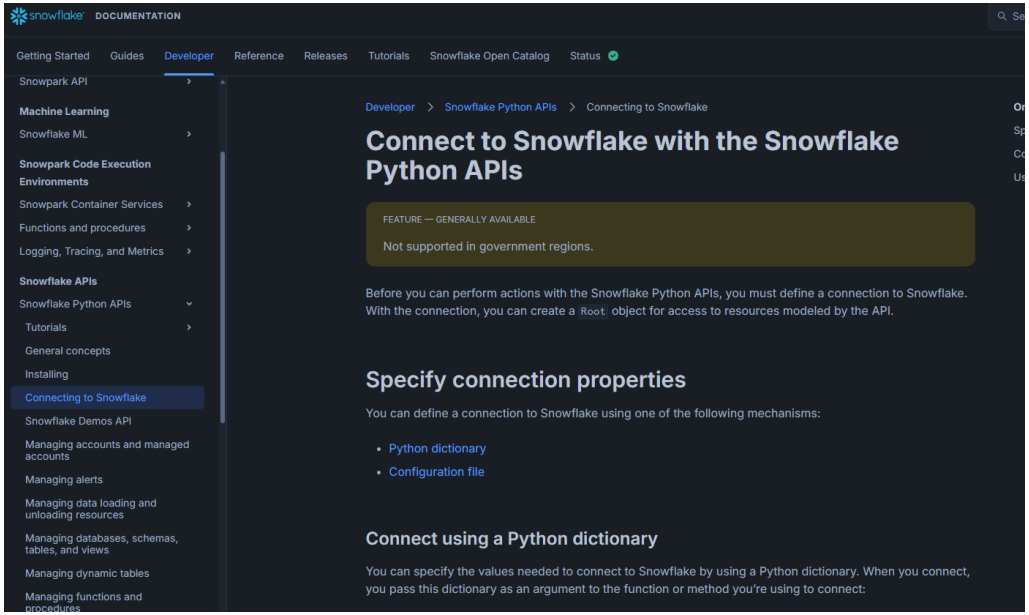
```
pip install snowflake-snowpark-python
```

Then the connector

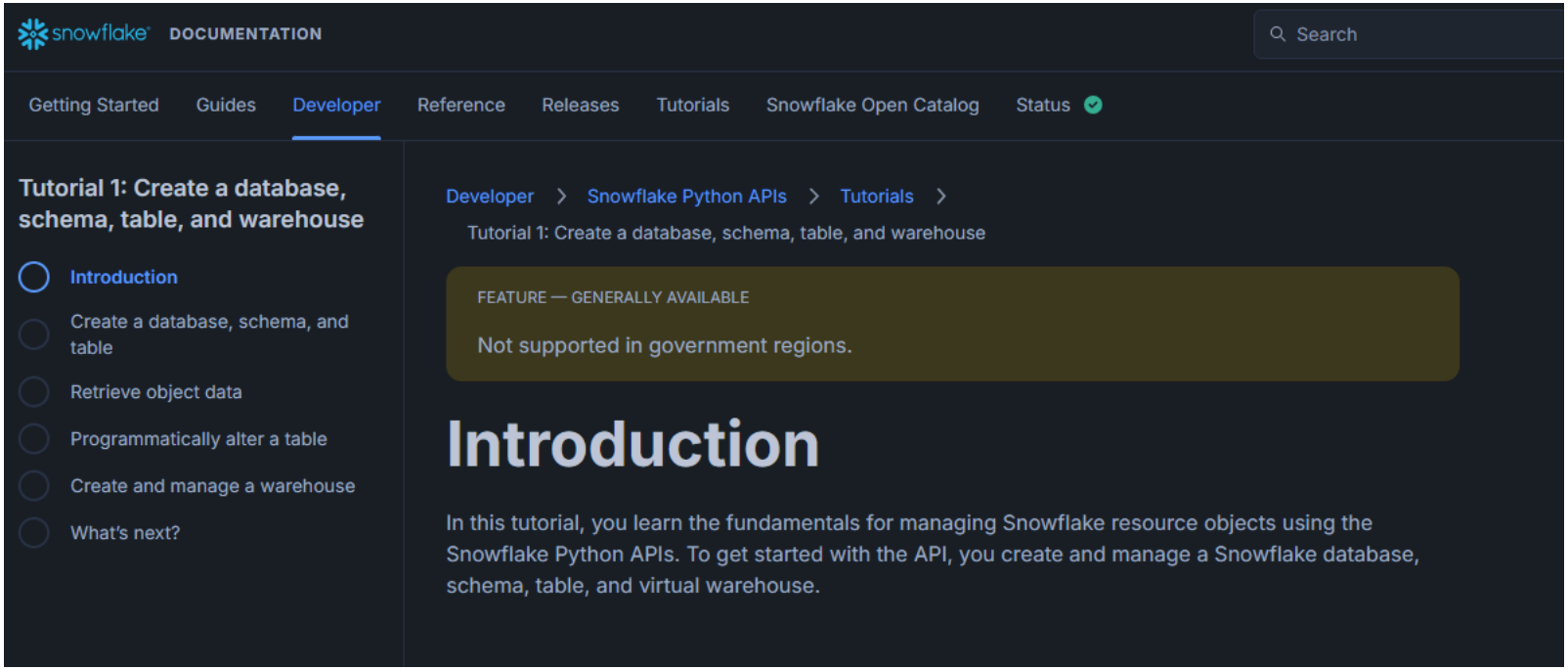
```
pip install "snowflake-connector-python[pandas]"
```

Click on the image below and follow the instructions to connect to your Snowflake Account.

Include screenshots and an explanation in your presentation.



Next, Click on the image below and follow the instructions to create a database and load your data into it.



You may need to use Copilot to find out how to load you Pandas data frame into snowflake.

Well done, now view your data in Snowflake, Include screenshots of your triumphant moment in your presentation.

	Unnamed: 0	title	Price	Mileage(miles)	Registration_Year	Previous Owners	Fuel type	Body type	Engin
1	0	SKODA Fabia	6900	70189	2016	3	Diesel	Hatchback	1.4L
2	1	Vauxhall Corsa	1495	88585	2008	4	Petrol	Hatchback	1.2L
3	3	MINI Hatch	2395	96731	2010	5	Petrol	Hatchback	1.4L
4	5	Hyundai Coupe	800	124198	2007	3	Petrol	Coupe	2.0L
5	9	Peugeot 207	1299	87000	2008	5	Diesel	Hatchback	1.6L
6	11	Vauxhall Insignia	2950	79000	2011	4	Petrol	Hatchback	1.8L
7	12	SKODA Fabia	6900	70189	2016	3	Diesel	Hatchback	1.4L
8	14	BMW 3 Series	19800	69690	2019	1	Petrol Plug-in Hybrid	Saloon	2.0L
9	17	Peugeot 207	999	111400	2007	2	Diesel	Hatchback	1.6L
10	19	Vauxhall Astra	1200	134000	2004	5	Petrol	Hatchback	1.8L
11	21	Ford Mondeo	1195	101000	2010	3	Diesel	Hatchback	1.8L
12	24	Ford KA	990	69000	2003	3	Petrol	Hatchback	1.3L
13	25	Vauxhall Insignia	2950	79000	2011	4	Petrol	Hatchback	1.8L
14	26	Mercedes-Benz A Class	15900	62402	2019	1	Petrol	Hatchback	1.3L
15	29	Citroen C4	3399	55000	2011	2	Petrol	Hatchback	1.6L
16	30	SKODA Fabia	1990	150000	2007	2	Diesel	Hatchback	1.9L
17	31	Ford Focus	1200	95000	2009	3	Diesel	Hatchback	1.8L
18	32	Mazda Mazda2	1490	135800	2008	3	Petrol	Hatchback	1.3L

Task 7: Create a Presentation

From the output of the steps above create a presentation slide deck (i.e. PowerPoint).

At the end of the presentation reflect on:

- What you've learnt,
- What issues you faced and how you solved them,
- why it is important or how could this be used in business,
- what you found difficult,
- and what you might do differently next time,

and finally, if you still want a career as a Data Engineer.

Task 8: Upload your challenge.

Create a ZIP file with your name in the filename, put the following in it a copy of:

Your presentation,

Upload it to the link that is in the questions section of the challenge email you were sent.

If you have any questions email them in and we will post answers to our blog below.

Please note: To be fair to all applicants we cannot answer individual questions.

Junior Tech Job Application FAQ's - Digital Native (dn-uk.com)