

## 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.

## 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](#)

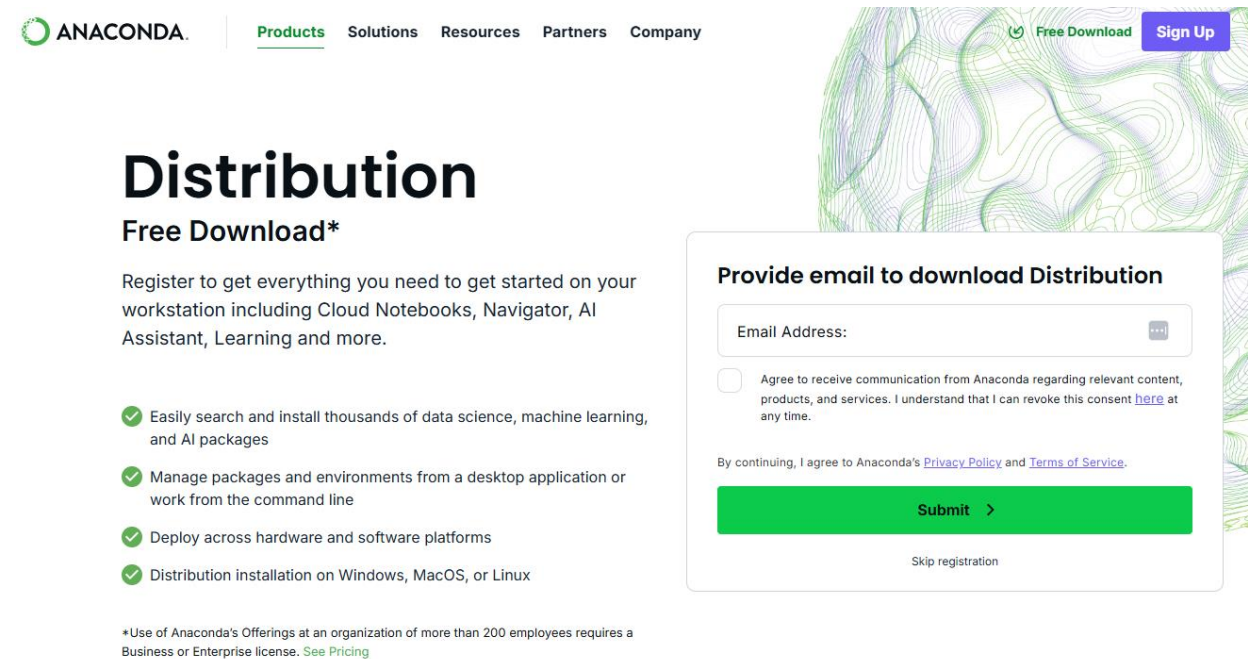
## Task 1: Download and Install Anaconda and Jupyter Notebooks

Jupyter Notebook is an easy way to start programming in Python as it is 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](#)



**ANACONDA.** | [Products](#) [Solutions](#) [Resources](#) [Partners](#) [Company](#)

## Distribution

### Free Download\*

Register to get everything you need to get started on your workstation including Cloud Notebooks, Navigator, AI Assistant, Learning and more.

- ✓ Easily search and install thousands of data science, machine learning, and AI packages
- ✓ Manage packages and environments from a desktop application or work from the command line
- ✓ Deploy across hardware and software platforms
- ✓ Distribution installation on Windows, MacOS, or Linux

\*Use of Anaconda's Offerings at an organization of more than 200 employees requires a Business or Enterprise license. [See Pricing](#)

### Provide email to download Distribution

Email Address:

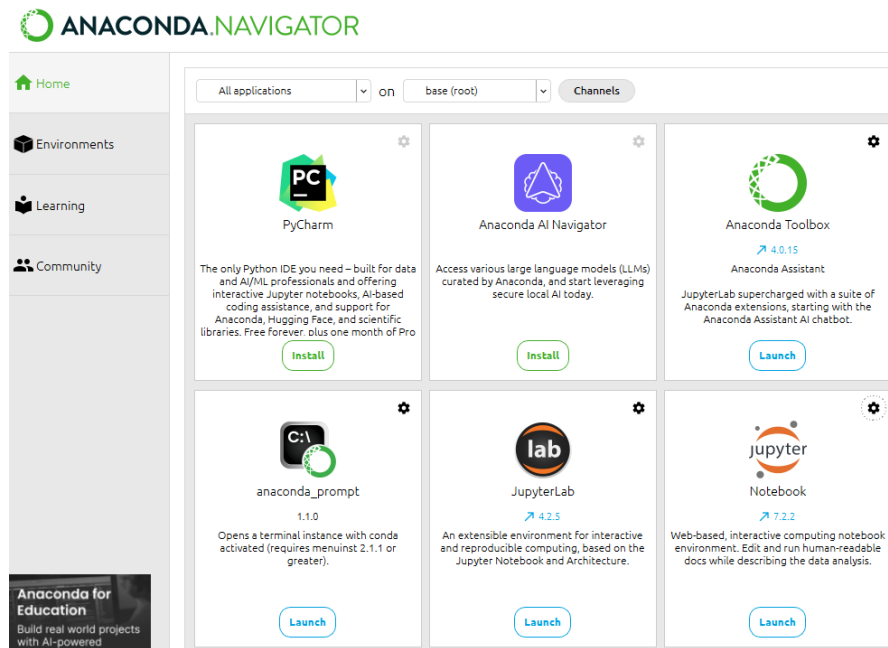
☐ Agree to receive communication from Anaconda regarding relevant content, products, and services. I understand that I can revoke this consent [here](#) at any time.

By continuing, I agree to Anaconda's [Privacy Policy](#) and [Terms of Service](#).

**Submit >**

[Skip registration](#)

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](https://www.kaggle.com),

kaggle

Competitions

Datasets

Models

Code

Discussions

Courses

...

Search

Sign In

Register

Click the image below and download the file:

Search

MUHAMMAD AWAIS TATYAB · UPDATED 2 YEARS AGO

39 Code Download

### Used Cars Prices in UK

Predict the price of used cars in UK

Data Card Code (21) Discussion (2) Suggestions (0)

#### About Dataset

Used Car Prices in UK Dataset is a comprehensive collection of automotive information extracted from the popular automotive marketplace website, autotrader.co.uk. This dataset comprises 3,685 data points, each representing a unique vehicle listing, and includes thirteen distinct features providing valuable insights into the world of automobiles. The feature names are:

- title
- Price : price of car in pounds
- Mileage(miles)
- Registration(year)
- Previous Owners
- Fuel Type
- Body Type
- Engine
- Gearbox
- Seats

#### Usability

10.00

#### License

CC0: Public Domain

#### Expected update frequency

Never

#### Tags

Data Analytics  
Data Visualization  
Exploratory Data Analysis  
Data Cleaning  
Regression

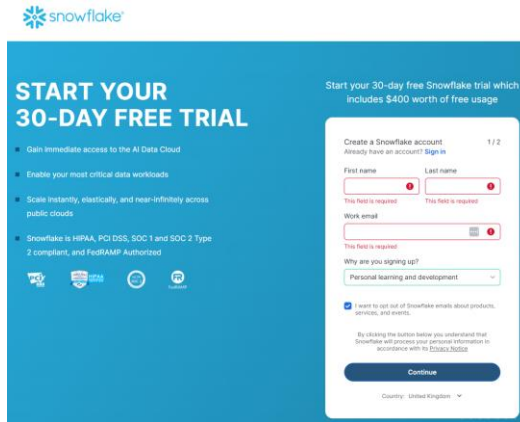
Alternatively follow the link here: [Used Cars Prices in UK](https://www.kaggle.com/datasets/muhammadawais1998/used-cars-prices-in-uk)

### Optional Stretch Challenge:

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

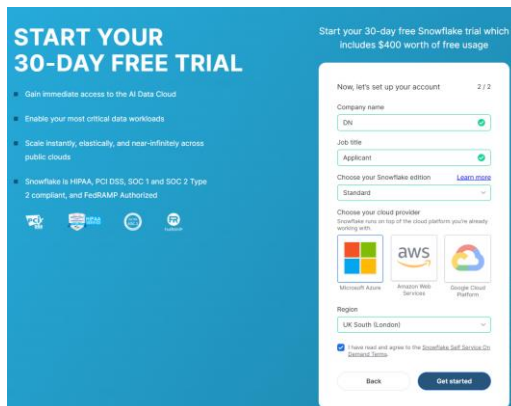
## Task 3: Sign up for a Free Snowflake Account.

Click on the image below, or [here](#) to create your own snowflake account



The image shows the Snowflake landing page for a 30-day free trial. The page has a blue header with the Snowflake logo and the text "START YOUR 30-DAY FREE TRIAL". Below the header, there are four bullet points: "Gain immediate access to the AI Data Cloud", "Enable your most critical data workloads", "Scale instantly, elastically, and near-infinity across public clouds", and "Snowflake is HIPAA, PCI DSS, SOC 1 and SOC 2 Type 2 compliant, and FedRAMP Authorized". To the right of the landing page is a "Create a Snowflake account" form. The form is titled "Create a Snowflake account" and has a progress indicator "1 / 2". It includes fields for "First name", "Last name", and "Work email", each with a red error message "This field is required". There is a dropdown menu for "Why are you signing up?" with the option "Personal learning and development" selected. A checkbox "I want to get out of Snowflake emails about products, services, and events." is checked. At the bottom, there is a "Continue" button and a "Country" dropdown menu set to "United Kingdom".

Set you location to the UK.







The image shows the same Snowflake landing page as before, but the form is now titled "Now, let's set up your account" and has a progress indicator "2 / 2". The form includes fields for "Company name" (filled with "DN") and "Job title" (filled with "Applicant"). There is a "Choose your Snowflake edition" dropdown menu with "Standard" selected. Below that is a "Choose your cloud provider" section with three options: "Microsoft Azure", "AWS", and "Google Cloud Platform". The "Region" dropdown menu is set to "UK South (London)". At the bottom, there is a checkbox "I have read and agree to the Snowflake Terms of Service" which is checked. There are "Back" and "Get started" buttons at the bottom of the form.

Select Python as your language

## START YOUR 30-DAY FREE TRIAL

- Gain immediate access to the AI Data Cloud
- Enable your most critical data workloads
- Scale instantly, elastically, and near-infinity across public clouds
- Snowflake is HIPAA, PCI DSS, SOC 1 and SOC 2 Type 2 compliant, and FedRAMP Authorized



Start your 30-day free Snowflake trial which includes \$400 worth of free usage

**Almost there...**

We're working on setting up your account. Help us better serve you by answering these questions.

Select your preferred language(s) to work in

☐ Scala

☐ Java

☐ SQL

☒ Python

☐ Other Tell us more





☐ I don't have coding experience

Skip

Continue

## START YOUR 30-DAY FREE TRIAL

- Gain immediate access to the AI Data Cloud
- Enable your most critical data workloads
- Scale instantly, elastically, and near-infinity across public clouds
- Snowflake is HIPAA, PCI DSS, SOC 1 and SOC 2 Type 2 compliant, and FedRAMP Authorized



Start your 30-day free Snowflake trial which includes \$400 worth of free usage

**Almost there...**

We're working on setting up your account. Help us better serve you by answering these questions.

What will you use Snowflake for?

☒ Load data, build a data pipeline or migrate an existing warehouse

☐ Build or distribute an application with Snowflake

☐ List or buy data from the Snowflake marketplace

☐ Use Snowflake as a data warehouse and data lake

☐ Use Snowflake Cortex AI to run LLMs, build RAG apps, and deploy ML models

☐ Build or train a machine learning model

☐ Run data analysis and build visualizations

Skip


Submit




Finally follow the link to the documentation,

## START YOUR 30-DAY FREE TRIAL

- Gain immediate access to the AI Data Cloud
- Enable your most critical data workloads
- Scale instantly, elastically, and near-infinity across public clouds
- Snowflake is HIPAA, PCI DSS, SOC 1 and SOC 2 Type 2 compliant, and FedRAMP Authorized






Start your 30-day free Snowflake trial which includes \$400 worth of free usage




### Check your inbox!

An email to activate your account has been sent to [tony@dn-uk.com](mailto:tony@dn-uk.com). It may take a few minutes to arrive.

Meanwhile, here are a few resources to check out:

-  [Get started with Snowflake documentation](#)
-  [Sign up for a free, instructor-led Virtual Hands-On Lab](#)
-  [Explore industry-specific user cases and walkthroughs in our Solutions Center](#)

WHAT IS  snowflake?

8 MINUTE DEMO

#### 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 are 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 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.

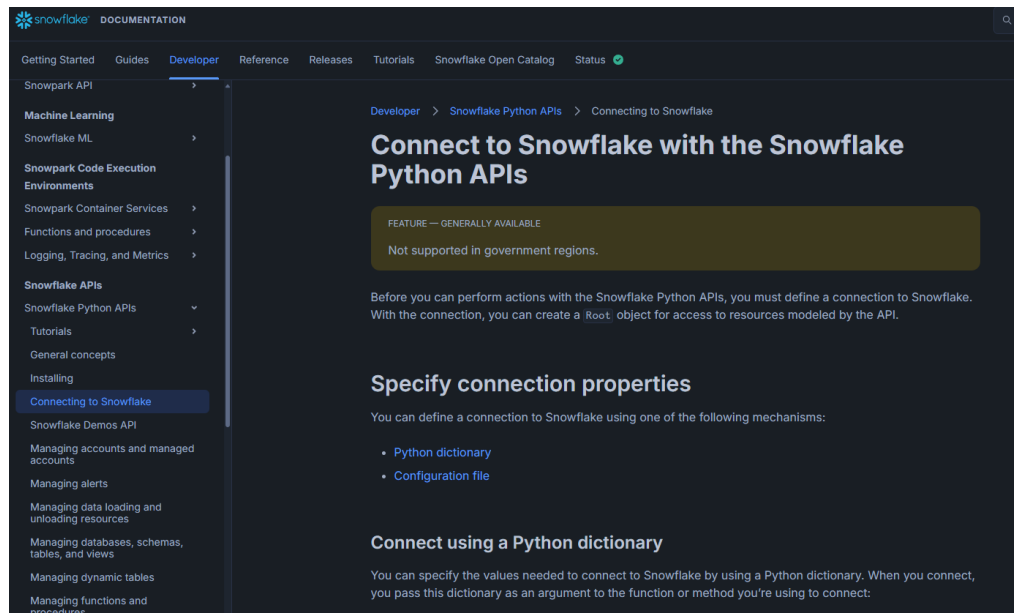
```
pip install snowflake-snowpark-python
```

Then the connector

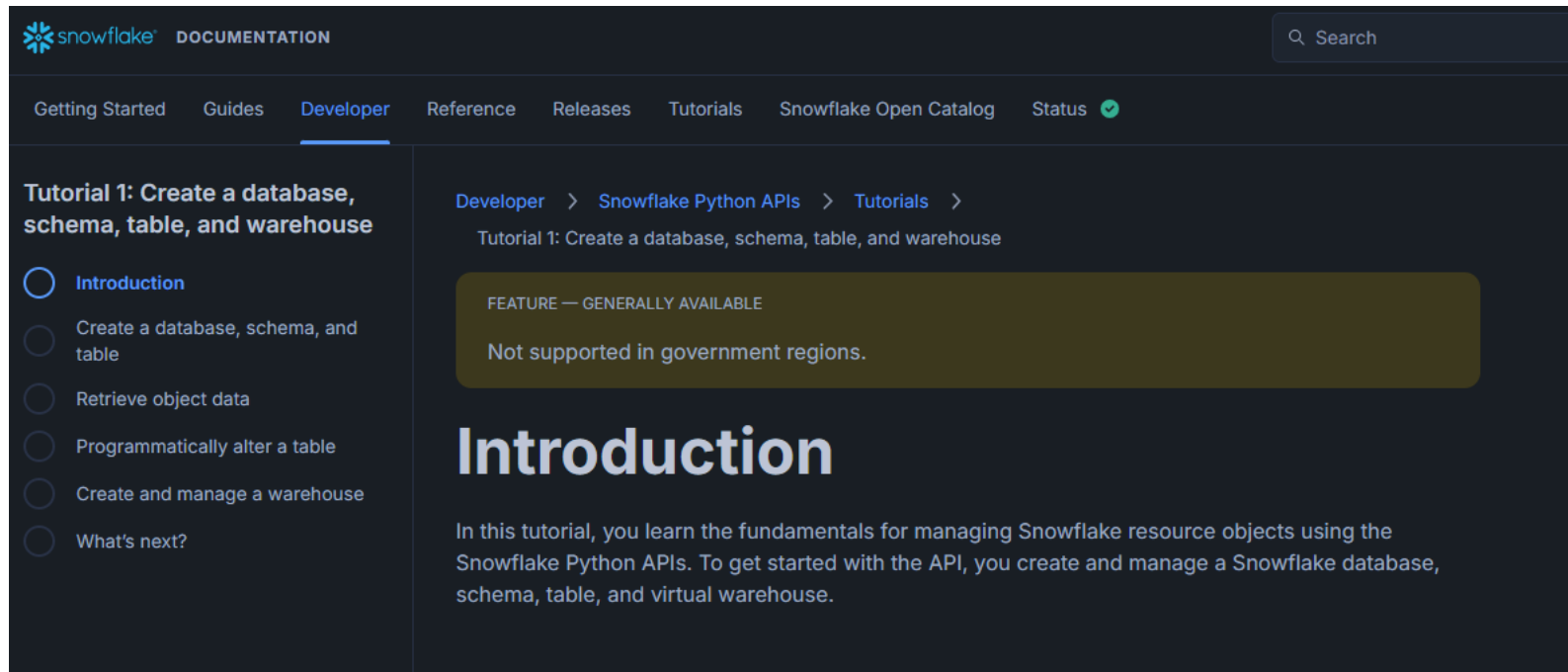
```
pip install "snowflake-connector-python[pandas]"
```

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.



The screenshot shows the Snowflake Documentation website. The top navigation bar includes the Snowflake logo, the word 'DOCUMENTATION', and a search bar. Below this is a secondary navigation bar with links: 'Getting Started', 'Guides', 'Developer' (which is highlighted with a blue underline), 'Reference', 'Releases', 'Tutorials', 'Snowflake Open Catalog', and 'Status' with a green checkmark icon.

On the left side of the 'Developer' section, there is a sidebar with the title 'Tutorial 1: Create a database, schema, table, and warehouse'. Below the title is a list of topics, each preceded by a radio button: 'Introduction' (selected), 'Create a database, schema, and table', 'Retrieve object data', 'Programmatically alter a table', 'Create and manage a warehouse', and 'What's next?'.

The main content area for the 'Introduction' topic shows a breadcrumb trail: 'Developer > Snowflake Python APIs > Tutorials > Tutorial 1: Create a database, schema, table, and warehouse'. Below the breadcrumb is a dark blue banner with the text 'FEATURE — GENERALLY AVAILABLE' and 'Not supported in government regions.'.

The main heading is 'Introduction'. The introductory text reads: 'In this tutorial, you learn the fundamentals for managing Snowflake resource objects using the Snowflake Python APIs. To get started with the API, you create and manage a Snowflake database, schema, table, and virtual warehouse.'

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



## 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 \(dn-uk.com\)](https://dn-uk.com/junior-tech-job-application-faq-s)