

# Goldman Sachs Challenge



## Introduction#

Data describing environments around the globe and documenting the behavior of financial markets has been collected for years. Modeling and machine learning allow us to spot trends and derive insights using the wealth of data available to us now. Our results can enable us to predict risk events, discover investment opportunities, and save entities and the environment from catastrophes.

## Description#

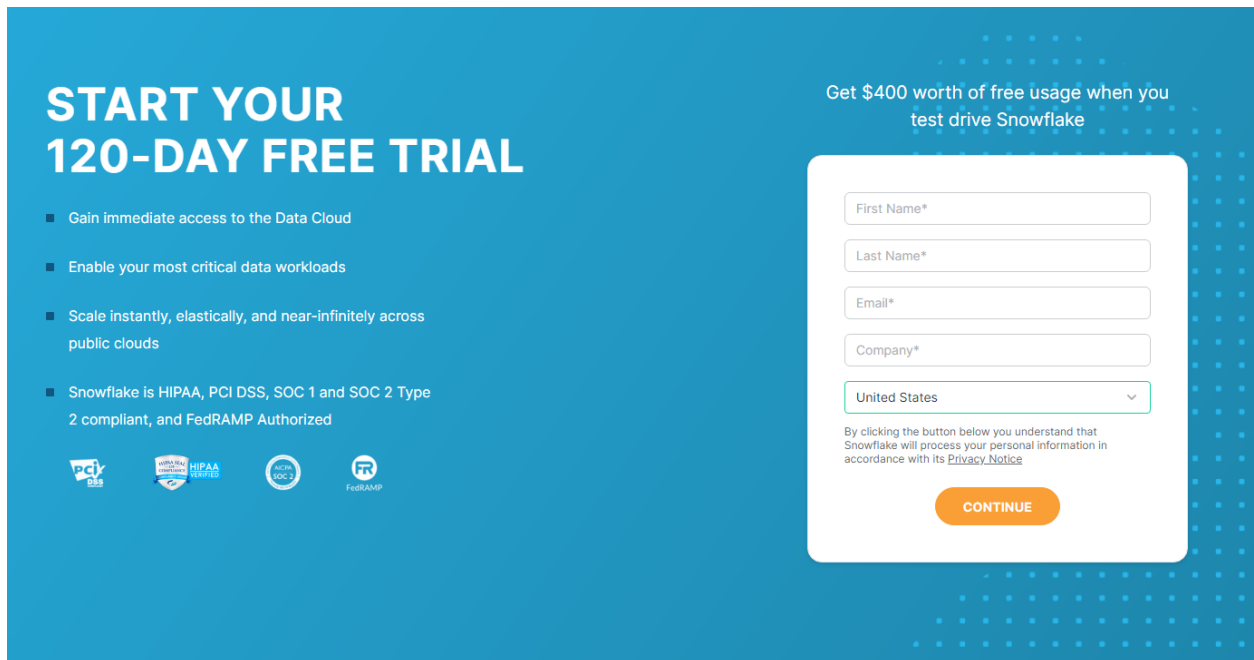
The final product of your analysis and modeling should document how the environment and the markets are tied together, and should provide further insight into how impactful the environment and markets are on one another.

Two Data Sources for the challenge will be sourced from the Snowflake Data Marketplace:

 <b>ZEPL</b>  <b>US Stock Market Data for Data Science</b>  US Stock and Index Daily Values from the Early 1980s Until Today	 <b>Knoema</b>  <b>Environment Data Atlas</b>  10+ public environment-related datasets from OECD, UN, World Bank, EDGAR, and other authoritative sources.
--	---

## Getting Started with Snowflake Data Platform#

1. Sign-Up for a Free Snowflake Account:  
<https://signup.snowflake.com/?trial=student>



The image shows the Snowflake sign-up page. On the left, there's a blue background with the text "START YOUR 120-DAY FREE TRIAL" in white. Below this, there are four bullet points: "Gain immediate access to the Data Cloud", "Enable your most critical data workloads", "Scale instantly, elastically, and near-infinately across public clouds", and "Snowflake is HIPAA, PCI DSS, SOC 1 and SOC 2 Type 2 compliant, and FedRAMP Authorized". Below the bullet points are logos for PCI DSS, HIPAA, SOC 1, and FedRAMP. On the right, there's a white box with a blue border containing a sign-up form. The form has fields for "First Name\*", "Last Name\*", "Email\*", "Company\*", and a dropdown menu for "United States". Below the form is a "CONTINUE" button. Above the form, it says "Get \$400 worth of free usage when you test drive Snowflake". Below the form, there's a small disclaimer: "By clicking the button below you understand that Snowflake will process your personal information in accordance with its [Privacy Notice](#)".

**START YOUR 120-DAY FREE TRIAL**

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

Get \$400 worth of free usage when you test drive Snowflake

First Name\*

Last Name\*

Email\*

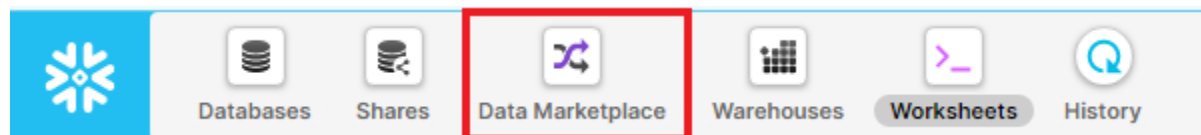
Company\*

United States

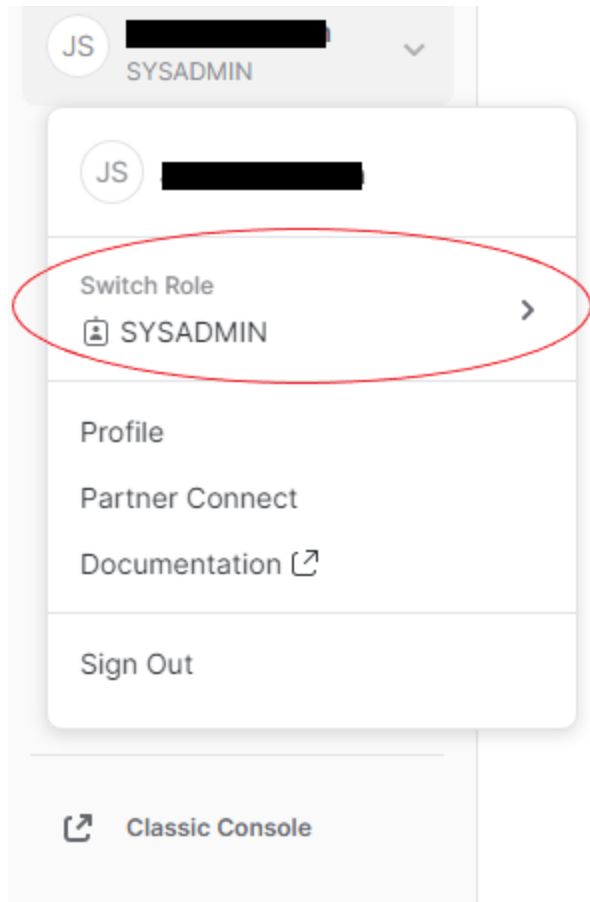
By clicking the button below you understand that Snowflake will process your personal information in accordance with its [Privacy Notice](#)

CONTINUE


2. After familiarizing yourself with the guided tour, navigate to the Data Marketplace



3. Set your "Role" to ACCOUNTADMIN




4. Search for and click on "Get Data" for the two data sources listed:



**ZEPL**

**US Stock Market Data for Data Science**

US Stock and Index Daily Values from the Early 1980s Until Today



**Knoema**

**Environment Data Atlas**

10+ public environment-related datasets from OECD, UN, World Bank, EDGAR, and other authoritative sources.

5. Navigate back to the Snowflake Console and begin working with the data through any of the supported methods Snowflake currently supports:

Snowsight: <https://docs.snowflake.com/en/user-guide/ui-snowsight-gs.html>

Snowflake Partner Connect:

<https://docs.snowflake.com/en/user-guide/ecosystem-partner-connect.html>

Snowflake Connectors: <https://docs.snowflake.com/en/user-guide/conns-drivers.html>

## Judging#

These are some guiding questions and listed areas of focus that we will take into account when judging your submission:

- Meaningful Insights from Data:
  - How do you develop metrics and insights which support your thesis?
- Dataset Usage/Integration:
  - Which methods do you use to mine, manipulate, and enhance the data?
- Data Normalization & Serialization:
  - When blending multiple datasets from different sources:
    - How do you manage complexity?
    - How do you maintain clean data?
    - How do you implement coherent transformations?
- Communication:
  - What will the final product look like?
  - How will you effectively convey your findings through your end-product?
- Documentation:
  - How will you shape your workflow and data usage documentation to clearly outline your methods while still meeting the timeline of the event?
- Automation:
  - How well will your models be able to accept new data and evolve to encourage future use?

## Prizes#

1st Place: Skullcandy Hesh Headphones (+ GS swag)

2nd Place: Goldman Sachs Swag

