BI Analyst home exercise

The goal of this exercise is to give you the opportunity to tackle new challenges, similar to the ones you'll face when coming onboard to Snyk, in order to empower Snyk to be able to make data-driven decisions throughout the organization.
In your answers please provide a detailed explanation to help us understand your thinking process. Please also including the queries you've used to fetch the data.

The dataset we'll be using is BigQuery , a Google-powered cloud hosting for very large datasets. BigQuery also hosts two interesting public data sets which we will explore in this exercise.
In order to access the BigQuery datasets you'll have to set up a free account.
The free access tier (1 TB scanning) will suffice for the purpose of this exercise.

GitHub is the world's largest and most popular platform for developers to host and collaborate their code. A project which hosts code on GitHub is called a "repository" (or repo in short).
In this part of the exercise, we will investigate GitHub's dataset.

a) How many repos from the sample_repos table have any programming languages data on them?

b) For the repos we found on section 1a:
What are the top 10 languages most commonly used?
How many repos use each language and what percent of the total repos does that represent?

c) Let's explore the sample_commits section.
A commit is a change of code that a developer is pushing to a code repository.

Find out the number of commits trend (YoY) for our top 10 most common languages:
i. Find out how many commits were added each year per each language in our top 10 languages.
ii. Build a simple excel/spreadsheet chart showing the trend for 3 languages of your choice.

Bonus Question
Stackoverflow is the largest website for developers professional Q&A.
Find one interesting insight from the 'stackoverflow' public dataset you think is worth mentioning.
In your answers please explain why you find this insight interesting.