Credit Risk Assessment Backend in GCP

Ali Ghorbani

Humberto Ribeiro de Souza

Credit Score Assessment in GCP

Business Problem

Architecture

Project objective

Implementation

Results

Implements endpoints for provide credit score in production environment

For U of T class 3760

Business Problem

- **Credit Default Risk as a Tool for Standardized Decision-Making**

  Credit risk tools are part of any financial institution to evaluate if money should be borrowed for all kinds of purposes. Some are purpose-built and more accurate for a given business case.

- **Credit Default Risk as an Inclusion Tool**

  Diverse initiatives across the globe have been designed and executed, some of them being even being awarded the Nobel Prize. In this case, the “Home Credit” financial institution has posted in Kaggle a problem that would provide credit to the unbanked population using alternative data such as telco and transactional information, to evaluate default likelihood. More can be found here.

Assumptions

01 Assumed that the Credit Home Company, which posted the Kaggle Challenge, has accepted the model to be promoted to production. The company has an aggressive growth business plan and could go tenfold in two years

02 Assumed that the team must select the best cloud provider for this need, considering that this is expected to be a low-profit and low-cost solution. Quality and simplicity are important, while response times are tolerable

03 The scope of this project is to provide an endpoint to Credit Home Company to query about credit default risk of new applicants.

04 Deprecated data for training models should be kept for audit purposes for 5 years. The model is expected to be updated fortnightly.

Cloud Technology Selection

- Project and Volumes

  Credit Home Corp expects to query between 600 - 1200 applications per day.

  The front-end will send a CSV containing a header and a row that belongs to the applicant.

- Security Risks

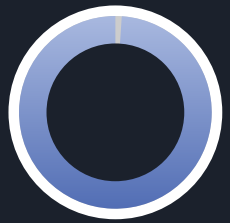
  The system handles both financial and personal data therefore the connection must be secured. Sniffing, Man-in-the-middle and classical attacks should be avoided.

  The company has sites in different states, has a team working from home and no VPN available at the moment.

On-Premises exclusion

Although the company has on-premise servers for their legacy Active Directory and related services, running models would be unsafe for the models are developed in Python 3.9 and Python libraries depend on operating systems library that needs to be patched.

Also, the solution would not scale according with the business plan without potential limitations to the business.



Cloud computing allows the application full-cycle to be managed without hardware and infrastructure concerns.

Cloud Provider Selection

The company has agreements with Google Cloud Platform and Amazon Web Services

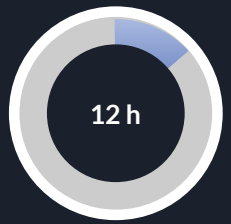
The estimated monthly costs for the PROD environment are available in the table below

Component	Expected Usage	AWS	GCP
Bucket per Gb	5 Gb inbound; 4 Gb moving within cloud	0.19 USD	**0.115 USD**
ML Engine per query	2 runs, notebook, 8 hours of nodes running	9.63 USD	**4.72 USD**
Database per Gb + access	30 Mbytes monthly AWS DynamoDB; GCP FireStore	0.30 USD	**Free**
Functions	500 ms of running code per req, 2 Gb of Ram, 1200 reqs per day	0.62 USD	**Free**
Total/Month		10.74 USD	**4.84 USD**

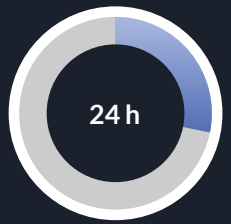
Architecture

Simplicity x Transparency

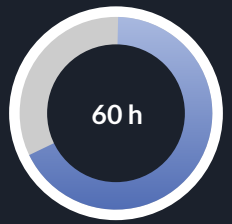
Different approaches were considered for building the backend. The simplest would be based on all resources available in Vertex AI. The implementation time would be the smallest. An serverless model has been chosen for audit and transparency for a reasonable amount of implementation time.



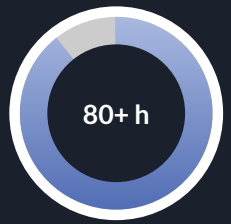
Vertex AI and API Endpoints, Firestore



Trained model available in Lambda, FireStore



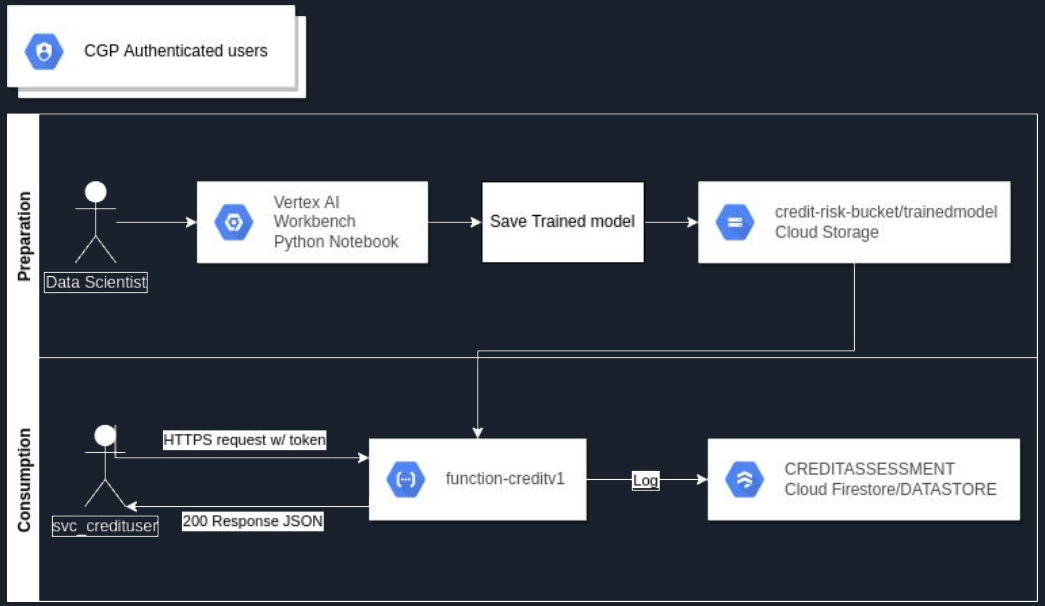
Cloud VMs, Network and access Management



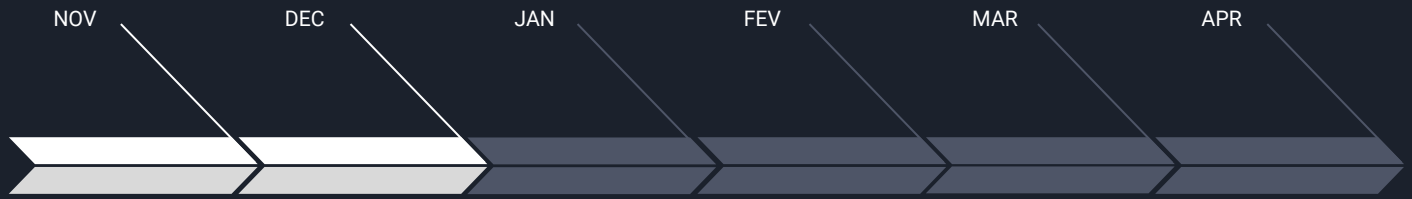
Acquire/setup hardware, OS, VMs, code

Architecture

The scope of this implementation is to expose the API endpoint for the Home Credit Company and allow assertion of the applicant credit score. The request must be recorded on a database for billing and auditing purposes. The model data is expected to be trained fortnightly and the previous data need to be stored so the model can be audited on a given point in time.



Project timeline - 2021/2022



Analysis & SOW

Discovery sessions, definition of done, selection of cloud provider, SOW

Implementation Phase 1

Model adjustment, environment creation, end-to-end tests, go-live

Review Phase 1

Weekly review; performance and security analysis. Preparation for phase 2

Phase 2 Analysis & SOW

Discovery sessions, definition of done, selection of cloud provider, SOW

Implementation Phase 2

Aim the automation of model data updates, audit review

Review Phase 2

Review of the ensemble of phases 1 and 2 weekly

Architecture

Justify selection of Cloud Functions over App engine or Cloud Run

- Event-driven solutions that extends to google are a good fit for cloud functions and we built function-based endpoint API therefore cloud function is the suitable choice.
- Less total cost of ownership and time-to-deploy
- App engine is suitable for many different services with a single application while cloud function preferred for individualized service.
- Cloud Functions-like services are found on other cloud providers and it is easily portable if a change of cloud provider is requested
- Cloud Run runs containers and is suitable choice to run in a stateless container. But, because of single query request of out end-pointAPI, we chose to use cloud function.

Architecture

Selection of Vertex AI Managed Notebooks over VM

- Maintain operating system (OS) and all software upgrade using VM is not core for the company nor its analysts
- Security risk of VM and continuously patching OS and ensure to keep it up to date
- Do not have native access to other GCP services and native integrations capabilities using VM. This can be arranged. However, the modules are frequently updated and it incurs in maintenance overhead.
- Not selected Auto ML to avoid platform dependency and maintain better visibility and easier explainability of the workflow

Architecture

Justify selection of FireStore/Datastore

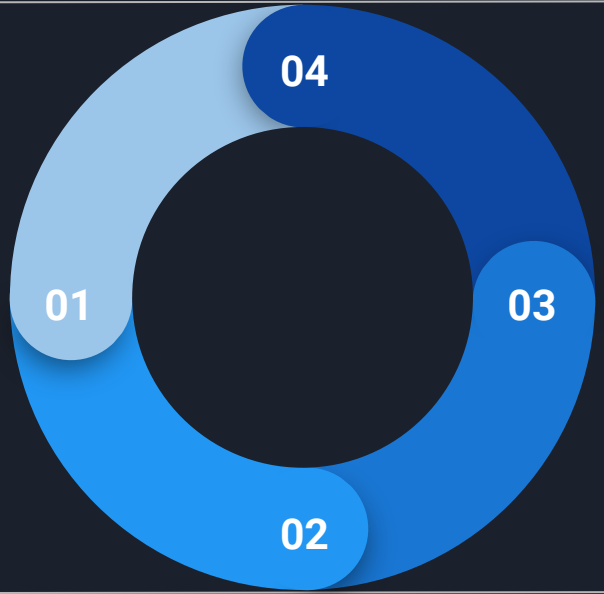
- Endpoint API needs a highly scalable document NoSQL database for small to medium size operational workloads
- It is cheaper than all listed contenders*
- Firestore automatically handles sharding and replication, with a highly available and durable database that scales automatically
- FireStore is built for automatic scaling, high performance, and ease of application development
- Firestore has key values, attributes and indices
- Firestore supports Atomicity, Consistency, Isolation, and Durability (ACID) transactions, SQL-like queries, indexes.

* https://sourceforge.net/software/compare/Amazon-DynamoDB-vs-Google-Cloud-Firestore-vs-Openredis-vs-Oracle-Database/

Cycle diagram

Update Model Fortnightly

A Data Analyst updates the model having the latest data at hand. Previous data is saved on bucket subfolders *



Audit

Review of results of a given applicant may be requested to run. The ensemble of information of usage and model data are necessary for the data analyst to run the audit*

API is enabled

The front-end system consumes it

Usage is recorded

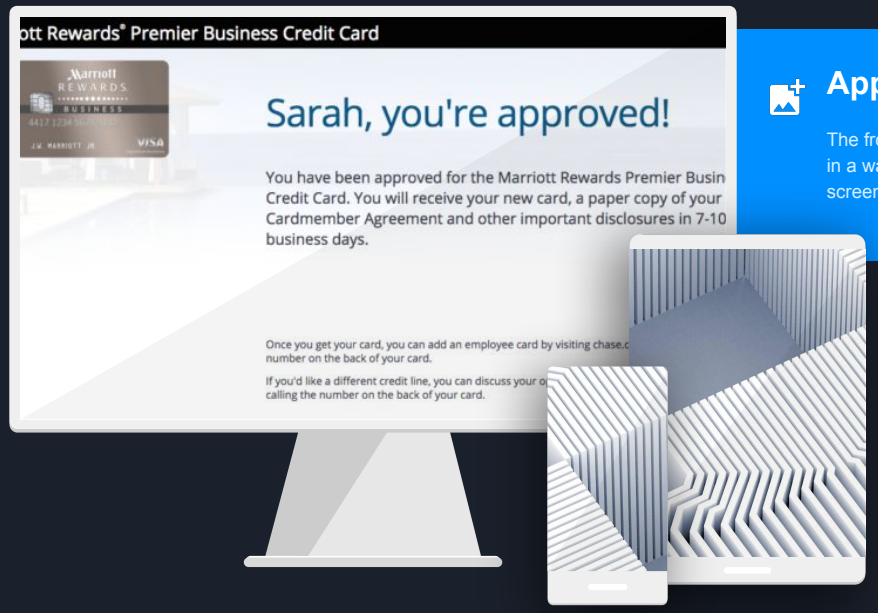
Usage is recorded for being queried by the front-end

* Automation of steps 1 and 4 are scope of another project

Introducing: Credit Risk Assessment System

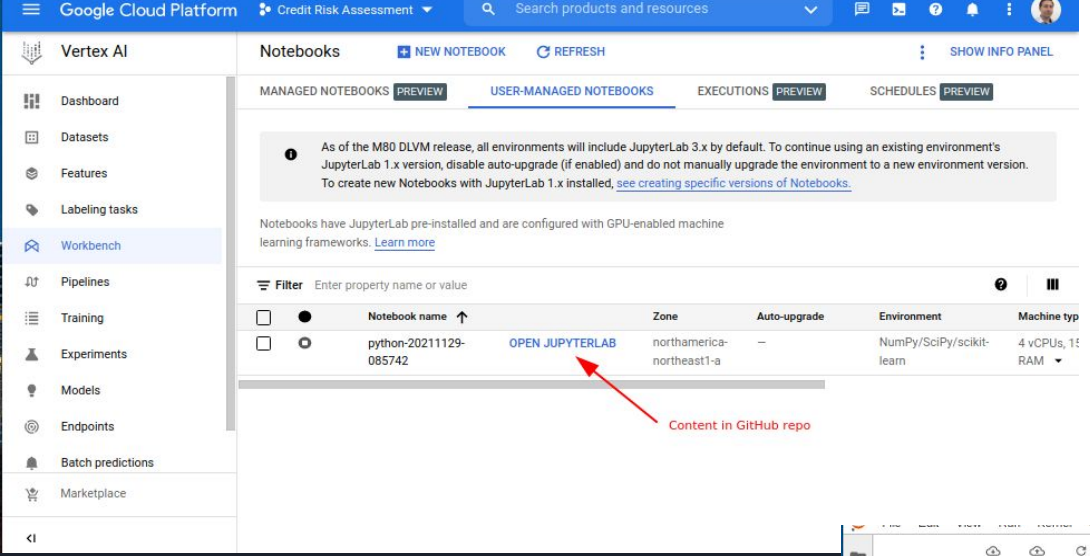
The endpoint is called from an application that runs on desktop, tablet or mobile

Access to the endpoint is open by default. It requires the service account svc_crediuser key to have access to it.

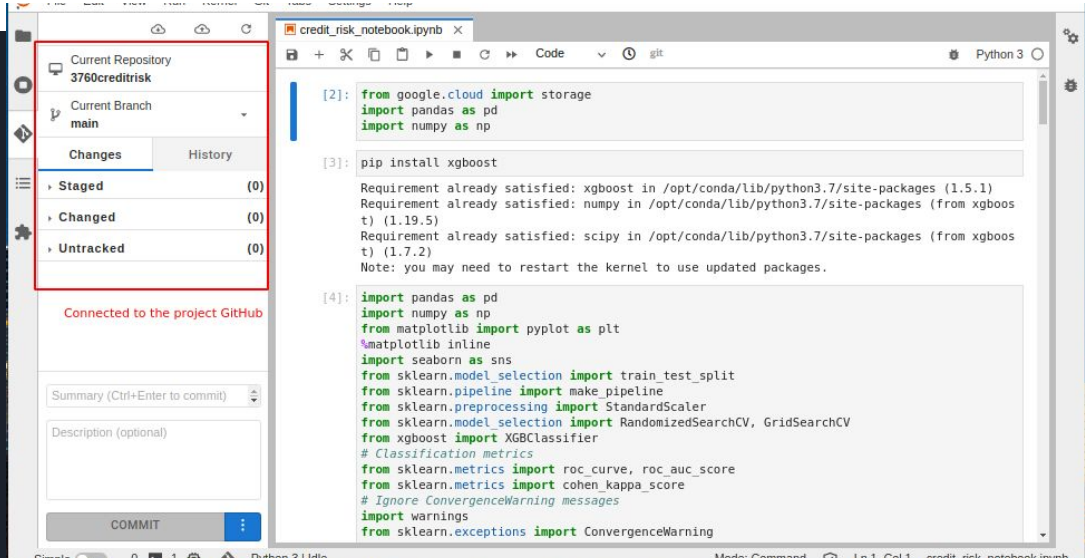


Approved/Reproved

The front-end system shows the status in a way the analyst can show the screen.



Vertex AI Notebook - Model building and exporting

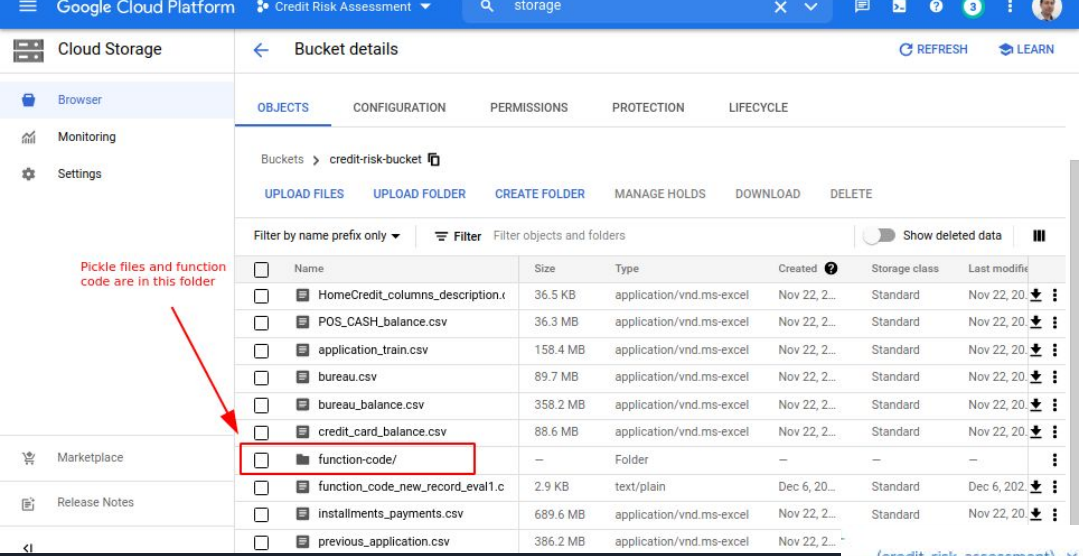


> Vertex AI - Notebook was used to adapt the model to work in the cloud.

> The train and test dataset have been cleaned-up and feature-engineered.

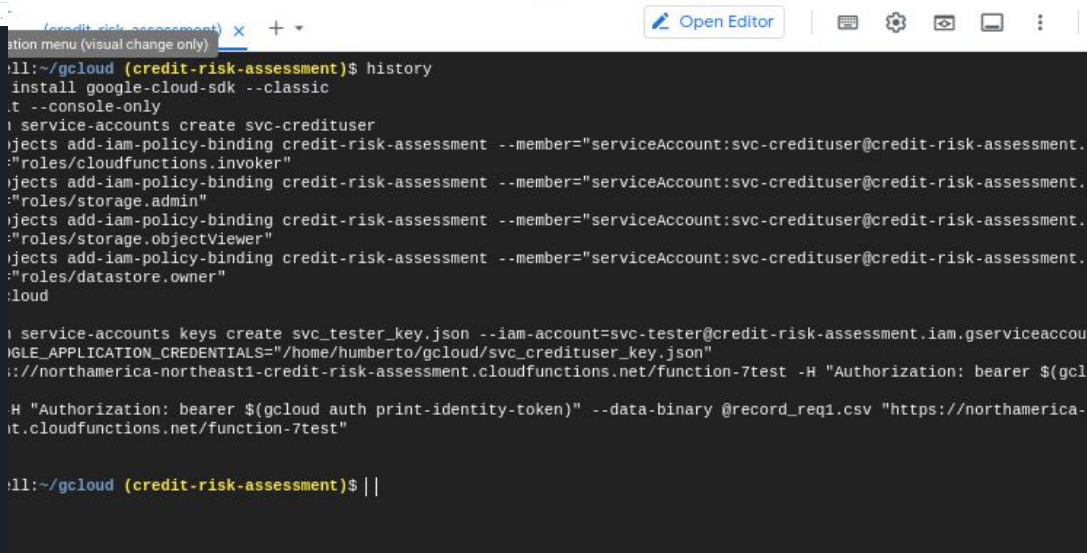
> The model uses XGBoost classifier using randomized search for hyperparameter optimization.

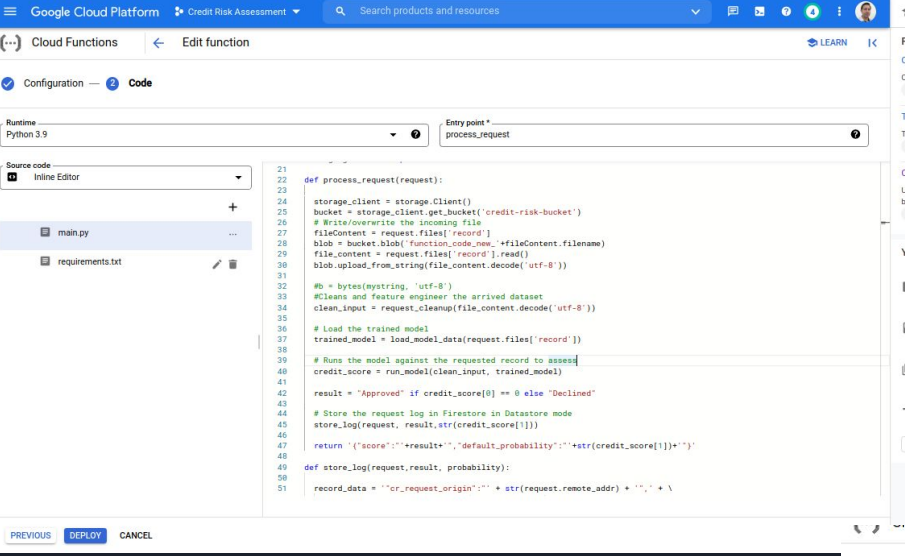
> The validation method was AUC for the dataset is imbalanced.



- Single bucket for the project: credit-risk-bucket
- Training model files for notebook are in the root
- Cloud functions-related code is sorted in functions-code folder
  - Trained model pickle
  - One-Hot encoded columns pickle
  - Zipped cloud functions code for automated deployment

Cloud Storage and service account IAM setup

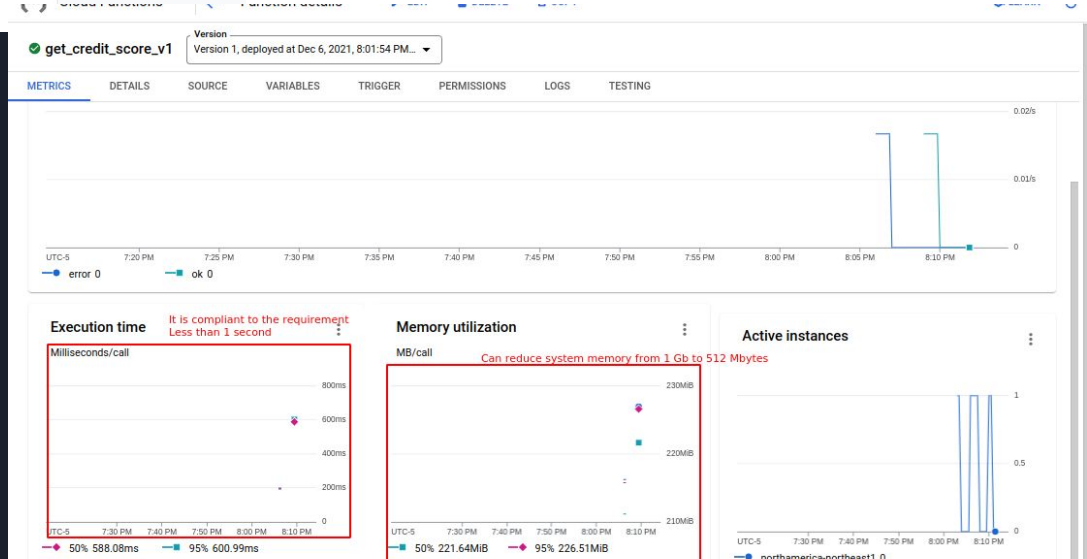


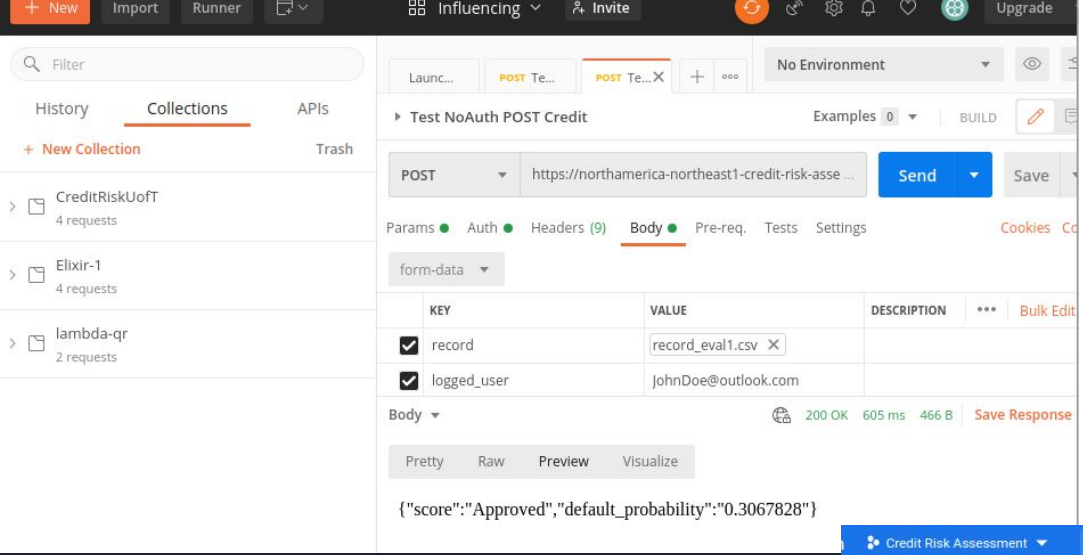


The function /get_credit_score_v1 calls

- The Cloud Storage for read and write operations
- Loads the trained model and reference record row
- Assess the record and returns the status
- Saves the transaction log in FireStore running as Datastore

Functions Setup





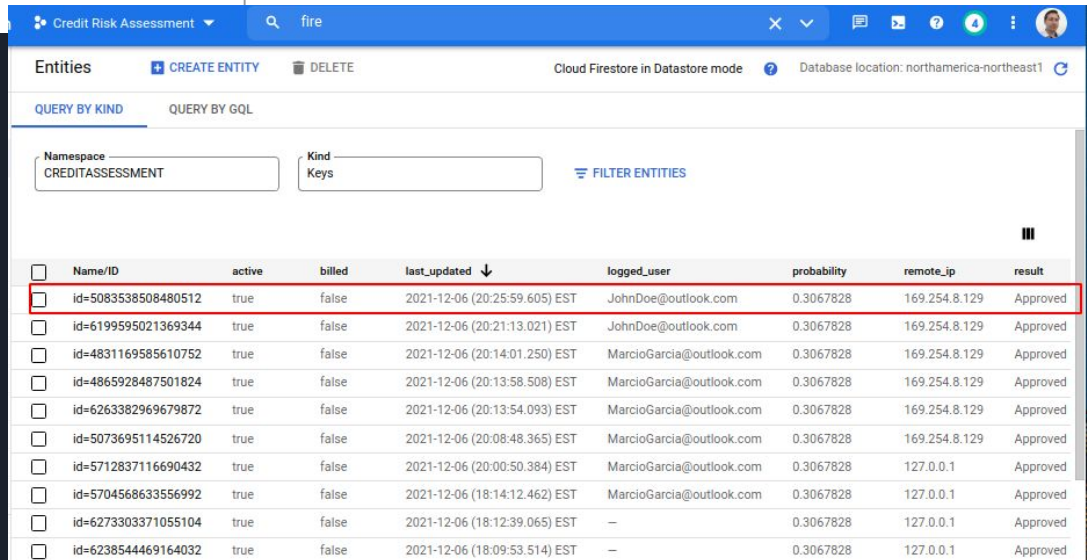
An authorized request can be made using Postman or curl.

All parameters are requested, along with the authorization key.

The result is a JSON record having

- Result: ['Approved','Declined']
- Default_probability in %

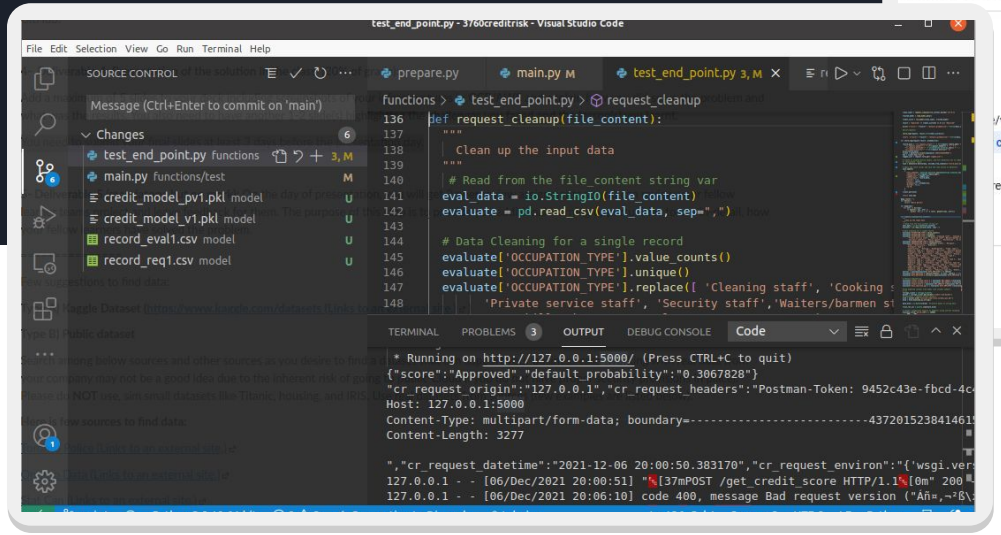
Validating with POSTMAN and FireStore

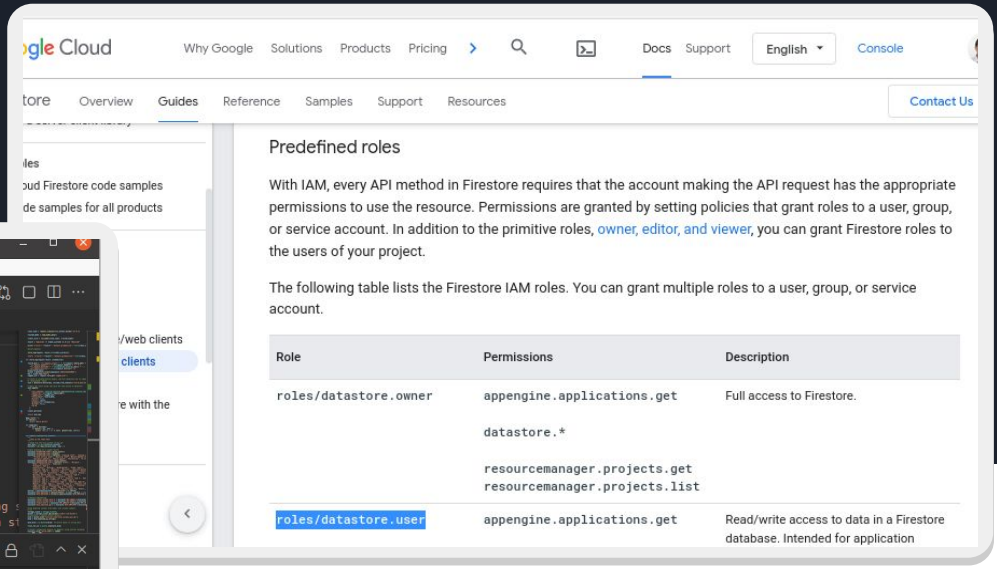


Name/ID	active	billed	last_updated	logged_user	probability	remote_ip	result
id=5083538508480512	true	false	2021-12-06 (20:25:59.605) EST	JohnDoe@outlook.com	0.3067828	169.254.8.129	Approved
id=6199595021369344	true	false	2021-12-06 (20:21:13.021) EST	JohnDoe@outlook.com	0.3067828	169.254.8.129	Approved
id=4831169585610752	true	false	2021-12-06 (20:14:01.250) EST	MarcioGarcia@outlook.com	0.3067828	169.254.8.129	Approved
id=4865928487501824	true	false	2021-12-06 (20:13:58.508) EST	MarcioGarcia@outlook.com	0.3067828	169.254.8.129	Approved
id=6263382969679872	true	false	2021-12-06 (20:13:54.093) EST	MarcioGarcia@outlook.com	0.3067828	169.254.8.129	Approved
id=5073695114526720	true	false	2021-12-06 (20:08:48.365) EST	MarcioGarcia@outlook.com	0.3067828	169.254.8.129	Approved
id=5712837116690432	true	false	2021-12-06 (20:00:50.384) EST	MarcioGarcia@outlook.com	0.3067828	127.0.0.1	Approved
id=5704568633556992	true	false	2021-12-06 (18:14:12.462) EST	MarcioGarcia@outlook.com	0.3067828	127.0.0.1	Approved
id=6273303371055104	true	false	2021-12-06 (18:12:39.065) EST	–	0.3067828	127.0.0.1	Approved
id=6238544469164032	true	false	2021-12-06 (18:09:53.514) EST	–	0.3067828	127.0.0.1	Approved

Top Challenges

1. Working with Cloud Functions specificities
2. Setting up IAM/Authorization roles without breaking everything
3. Choosing tools for a forecast of exponential growth





Thank you!

Ali Ghorbani

Humberto Ribeiro de Sousa

2021-12-07