Techionista Academy

Microsoft Azure Certified Data & AI Track

Techionista Holidays Business Case

September 2022

Table of Contents

Techionista Holiday - Business Case

About the Assignment

What will you need to do during this business case? Find out during this section of the guide!

The Assignment

Learn more about your challenge during this business case.

Case Study

Going on vacation is lovely, but half the fun is in planning and anticipating your trip. Scrolling and sifting through the endless options for days, taking into account all of these different factors to organize the perfect getaway. Sound familiar?

Finding suitable accommodation is an important part of planning a vacation and that's what this business case is all about! You'll be working with an extensive dataset from Techionista Holidays. Among other information, it contains reviews from their users after their stay at a hotel they booked through the Techionista Holidays website. These reviews date from 2015 to 2017 and contain lots of information on how travelers have experienced their stay in various luxurious hotels across Europe. Techionista Holidays managers would love to gain more insight into the different types of travelers and their feedback. These valuable insights can help them improve Techionista's Holidays website and their capability of catering to their users needs.

Your Challenge

These reviews consist of free text fields and that's the challenging bit. While these text fields contain a lot of information, it's hard to make aggregated conclusions since the data is qualitative and not numerical.

Reviewing every written statement about the hotels would be a fool's errand...

Azure Cognitive Services

Luckily, there is an easier solution! With Power BI, it's possible to make use of various functions from Azure Cognitive Services. During this business case, you will make use of a sentiment analysis function. This function, which is based on an AI model, determines the sentiment score of a text and can extract key words from it. A great way to quickly gain insight into thousands of reviews.

In short

For this assignment, you will work in groups of four to five people. The first step will be to connect to the data source and clean the data. Then, you will have to set up the Azure Cognitive Services in Azure and run the function in Power BI. The last steps of the assignment will be to build a report or dashboard and present it. Make sure you prepare a short presentation of five minutes.

Important: aside from the end result, being a report/dashboard, we also like to receive a detailed step-by-step plan of the steps you've taken while working on the business case. So make sure to keep track of your steps as soon as you start working on the business case!

Your Goal

This assignment is aimed at preparing you for your future job by providing you with hands-on experience working with Power BI. It includes parts of the profession of Data Engineer as well as Data Analyst. You will collect the data, clean it, and build a report or dashboard.

Note: good to realize is that not everything is laid out in front of you but during this assignment you also have to do your own research and apply it to your assignment.

Disclaimer

The dataset that you will be working with is publicly available to everyone. Please be noted that the data is originally owned by Booking.com. Unfortunately, we can't use that name, that's why we transformed it to Techionista Holidays.

Techionista Holiday - Business Case

Getting Started

The first steps to get you started on your business case.

Getting Started

Learn more about how to get started for the business case.

Step One: Email Account

Create a new Outlook or Hotmail email account. And write down this new email account and password!

Note: an Azure pass can only be activated using a Microsoft email address (Hotmail or Outlook) and only one pass can be activated per email address.

Step Two: Azure Pass

Per group you will receive one Azure Pass. An Azure pass is a token that gives you full access to the Microsoft Cloud for a given amount of time and money. The Azure pass is valid for 30 days or until you run out of credits.

Activate your Azure pass here.

Note: To avoid issues, use an incognito/InPrivate window. Ensure the Azure pass page isn't linked to a previous or wrong Outlook/Hotmail account. If so, make sure to log out and use your newly created email address before you activate the new Azure pass.

How do I know how much credit is left on my Azure Pass?
You can check out your Azure Pass balance using the following website here.

Connect to the Data Source

Open Power BI Desktop and choose Get Data.

The dataset is stored in Azure Blob Storage (see screenshot below).

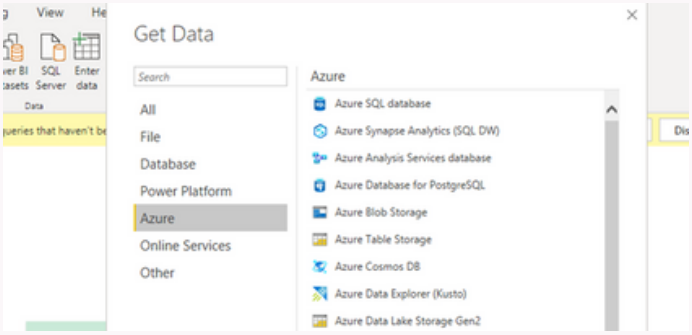
To connect to Azure Blob Storage, you need to fill in the following details (Do this first in a notepad so you can make one sentence out of it, then copy the whole sentence into Power BI):

URL:
https://pl300businesscase.blob.core.windows.net/pl300businesscase

KEY:
++LZ+iItA/PflyqSAXu+2yz7O3PVwDWrClhle8f1xcHJb1rdkhJi7jMXYvO/jaPJRkcCdh9/IpxvUOasqDEnaQ==

Note: make sure to transform the dataset first, do not load it yet!

Note: for this business case you will work with Power BI. Please do not make use of the Power BI Pro trial, since you don't need a Pro license for this business case. You will claim the Power BI Pro trial later on in this track.



Techionista Holiday - Business Case

The Dataset

Learn more about the dataset in this section of the guide.

The Dataset

Learn more about how to get started for the business case.

The Dataset

This dataset contains 515,000 customer reviews and scoring of 1493 luxury hotels across Europe.

It contains 17 fields. The description of each field is as below:

- Hotel_Address: Address of hotel.
- Review_Date: Date when reviewer posted the corresponding review.
- Average_Score: Average score of the hotel, calculated based on the latest comment in the last year.
- Hotel_Name: Name of hotel
- Reviewer_Nationality: Nationality of reviewer
- Negative_Review: Negative review the reviewer gave to the hotel. If the reviewer does not give the negative review, then it should be: 'No Negative'
- Review_Total_Negative_Word_Counts: Total number of words in the negative review.
- Positive_Review: Positive review the reviewer gave to the hotel. If the reviewer does not give the positive review, then it should be: 'No Positive'
- Review_Total_Positive_Word_Counts: Total number of words in the positive review.
- Reviewer_Score: Score the reviewer has given to the hotel, based on his/her experience
- Total_Number_of_Reviews_Reviewer_Has_Given: Number of Reviews the reviewer has given in the past.
- Total_Number_of_Reviews: Total number of valid reviews the hotel has.
- Tags: Tags reviewers used in their review.
- days_since_review: Duration between the review date and scrape date (the scrape date being the date the data was scraped of the Booking.com website, which is about 4 years ago)
- Additional_Number_of_Scoring: There are also some guests who just made a scoring on the service rather than a review. This number indicates how many of those valid scores a hotel has that don't include a review
- lat: Latitude of the hotel
- lng: Longtitude of the hotel

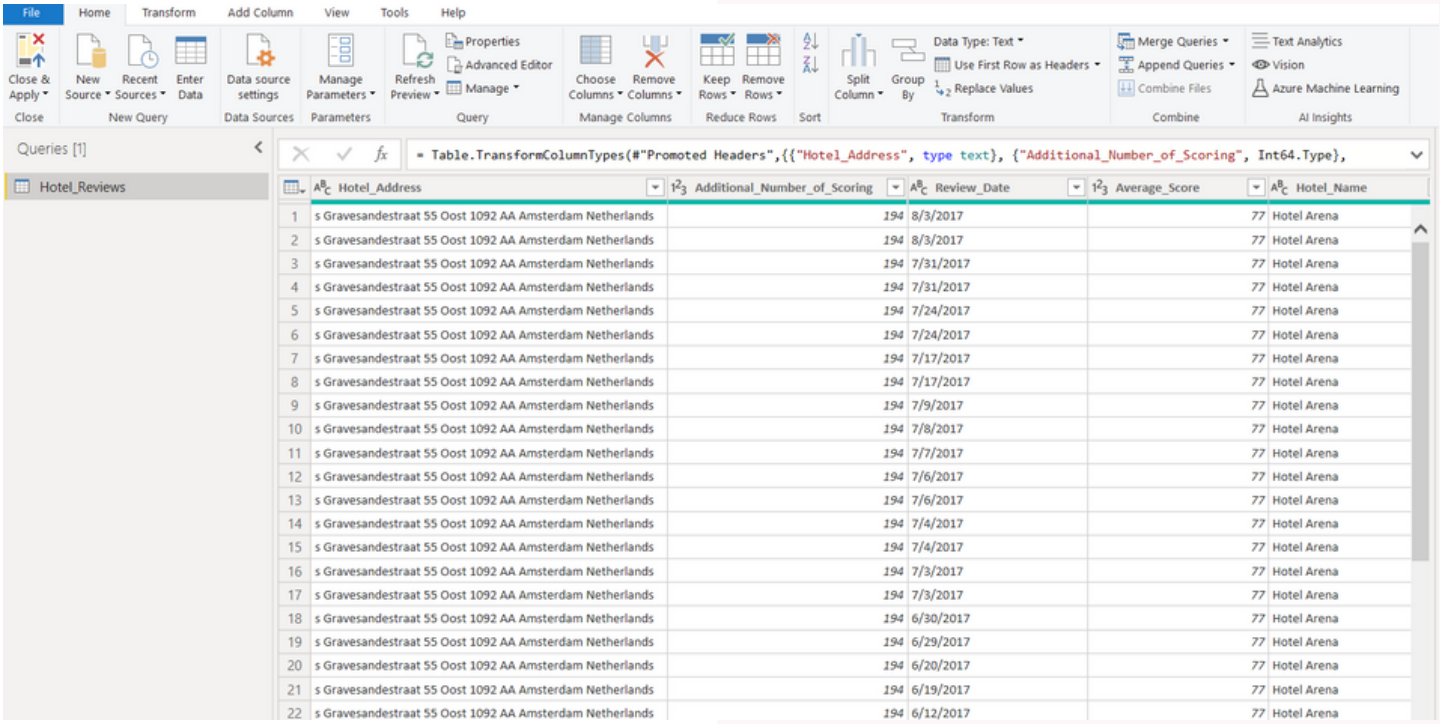
Transforming and Cleaning

Learn more about how to get started for the business case.

Transforming and Cleaning the Data

In **PowerQuery**, open the content of the file (see screenshot below).

- Check for errors in the dataset. Remove the errors.
- Check for incomplete data (Null values), if necessary, change Null values to Unknown.
- Do you see any other incomplete data? If so, change it as well.
- Make sure the headers have clear names.
- Format the columns correctly.
- Split the contents of one column into multiple columns in case you'd like to work with specific parts of the column content.
- Select Keep Rows, keep top 75 rows! **We will explain the reasoning behind this step later on.** **See page 15.**



Setup the Cognitive Services in Azure

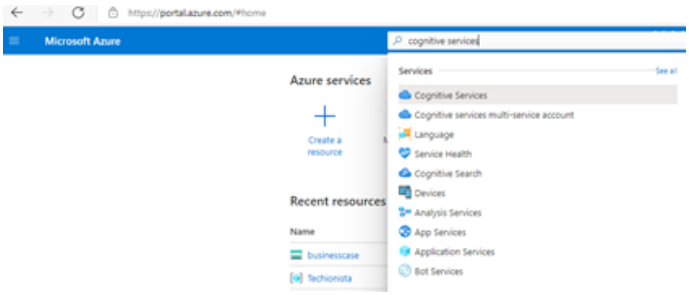
Learn more about how to get started for the business case.

Cognitive Services

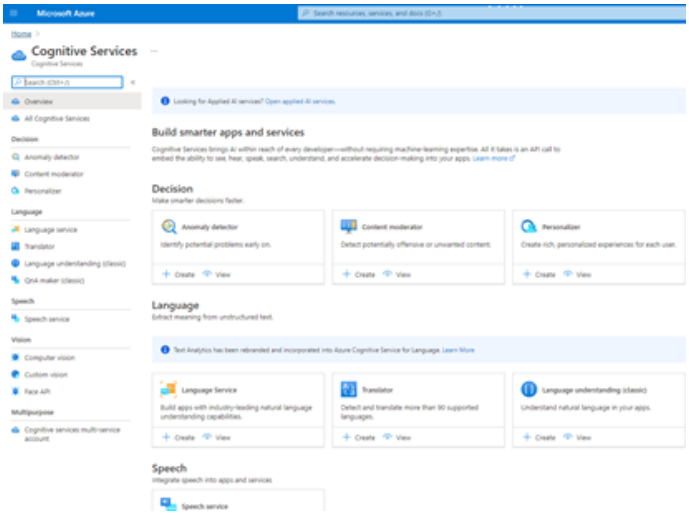
Cognitive Service runs in Azure. We first need to set this up before we can connect to it in Power BI.

- Go to this website.
- Above in the blue search area type in: Cognitive Services.
- Choose the first option: Cognitive Services (see screenshot 1).
- You will see screenshot 2. Now you can pick which cognitive service you want to use and click on '+ Create'.
- For this example, I will show you what happens if you click on '+ Create' at Cognitive Services 'Language Service' (see screenshot 3).

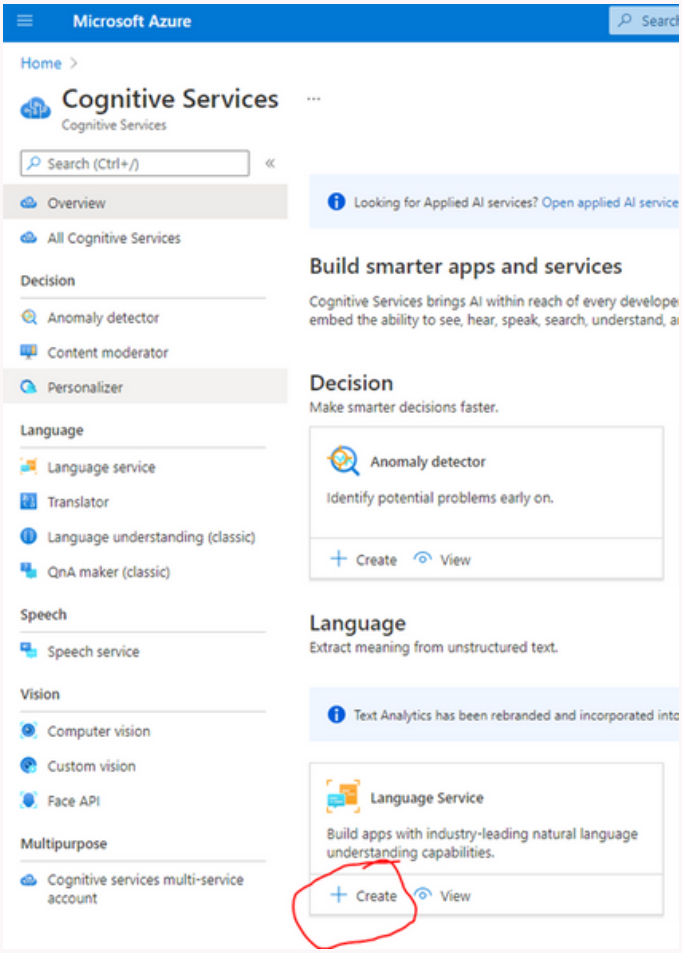
1



2



3



Setup the Cognitive Services in Azure

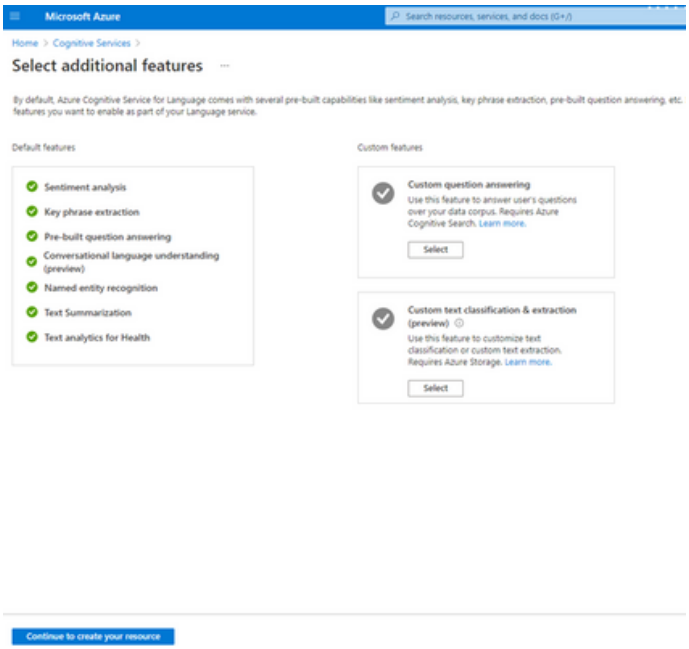
Learn more about how to get started for the business case.

- Choose a custom feature or leave it open (this depends on your plan and what you need) and click on 'Continue to create your resource' (see screenshot 4).

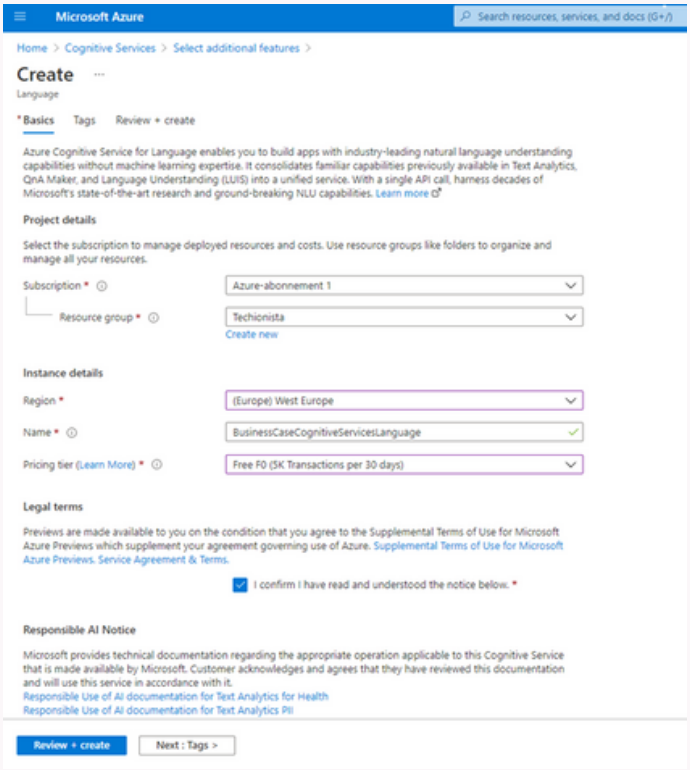
- To create your Cognitive Services Language: - Select your resource group (this is the one you made earlier);
  - Select the region: (Europe) West Europe;
  - Fill in a name for your Cognitive Services Language Service (come up with your own name);
  - Select the pricing tier: Pick the Standard tier!
  - Check the two boxes;
    - Click on Review + Create;
    - Click on Create.

- If you have done all this, then wait a few minutes while your deployment is in progress. Go to the resource by clicking on the blue button, see screenshot 6.

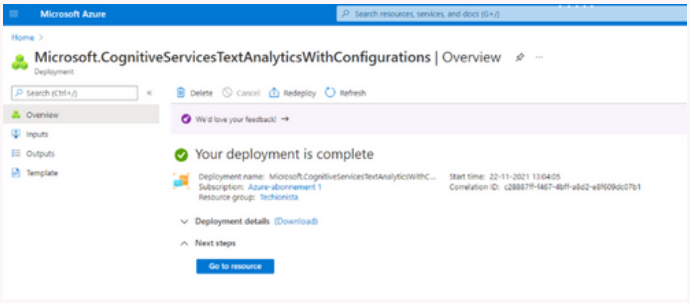
4



5



6

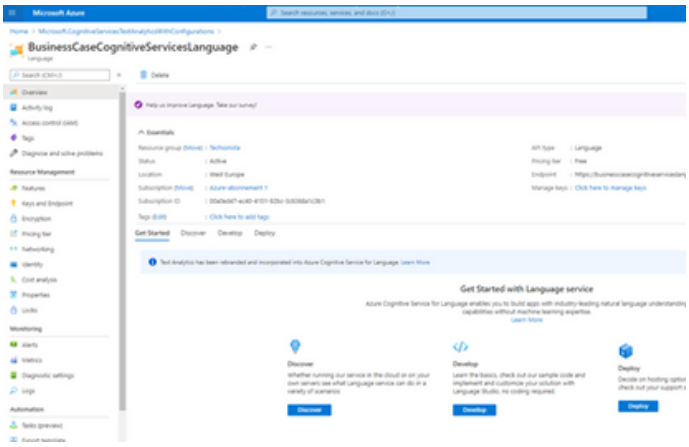


Setup the Cognitive Services in Azure

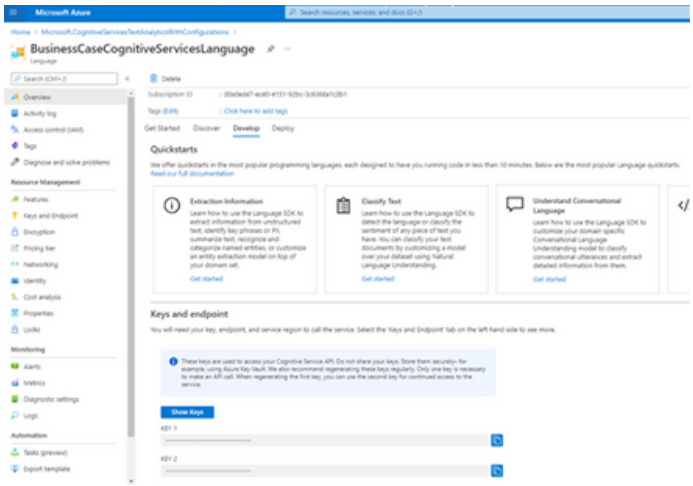
Learn more about how to get started for the business case.

- Now you see screenshot 7.
- If you go to ‘develop’ you can pick up your key and endpoint. You will need these in your code in Power BI, so write them down! If you click on the blue button ‘show keys’ you will see them. Scroll down for your endpoint (see screenshot 8).
- Go to deploy and click on ‘Get Started’ (see screenshot 9).
- After you have clicked on ‘Get Started’, you see screenshot 10.

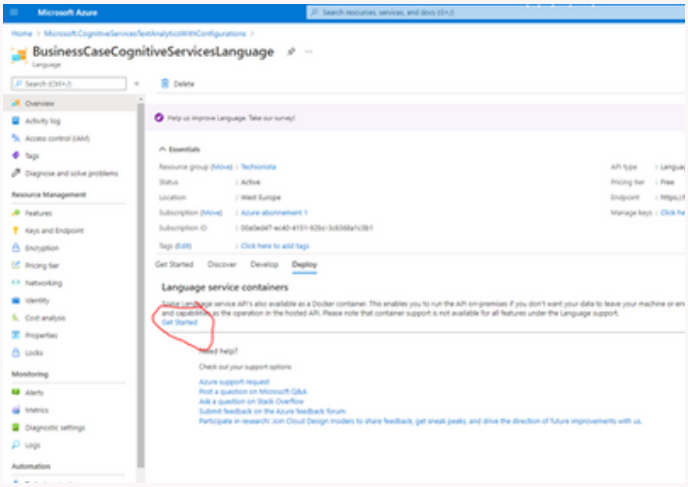
7



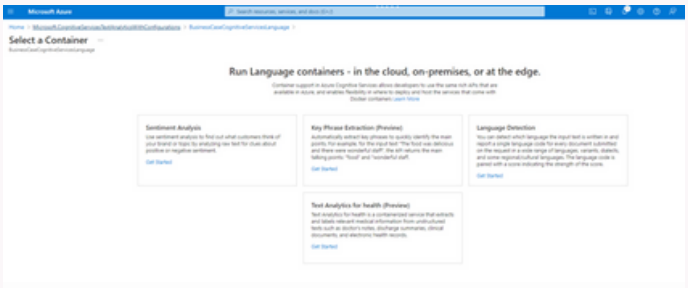
8



9



10

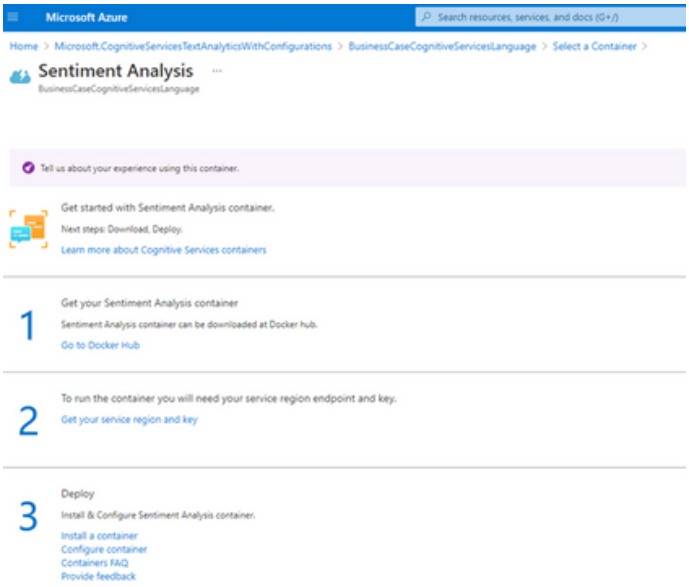


Setup the Cognitive Services in Azure

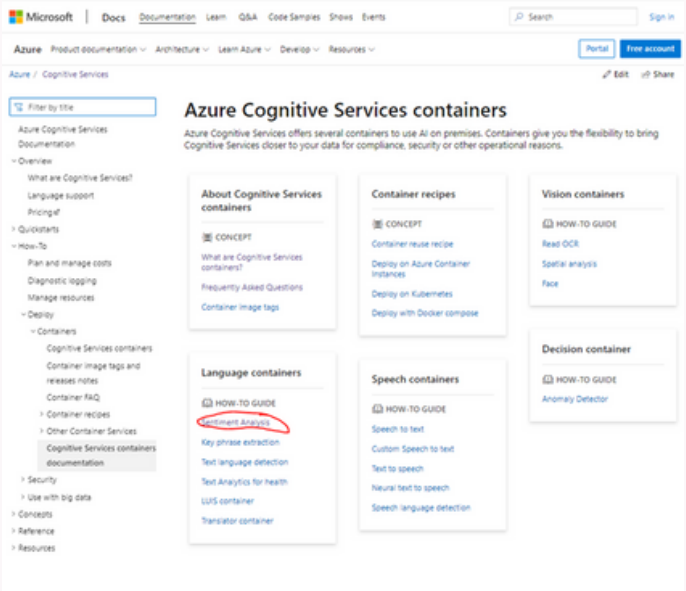
Learn more about how to get started for the business case.

- Click on the language container you want to use. In this example, I have chosen the 'Sentiment Analysis' (see screenshot 11).
- Good to know:
  - You don't need the Docker Hub;
  - You already have the key and endpoint;
  - You can now continue on the next page of this guide (step: Create the Power Query function in the Query Editor) or you can follow the next steps.
- You can go to: Learn more about Cognitive Services containers.
- Under Language containers, how-to guide, click on Sentiment Analysis (see screenshot 12).
- Here you go to 'Tutorials' and click on 'extract key phrases from text stored in Power BI' (see screenshot 13).
- Now you see a tutorial about extract key phrases. You can use this for your dashboard/report in Power BI.

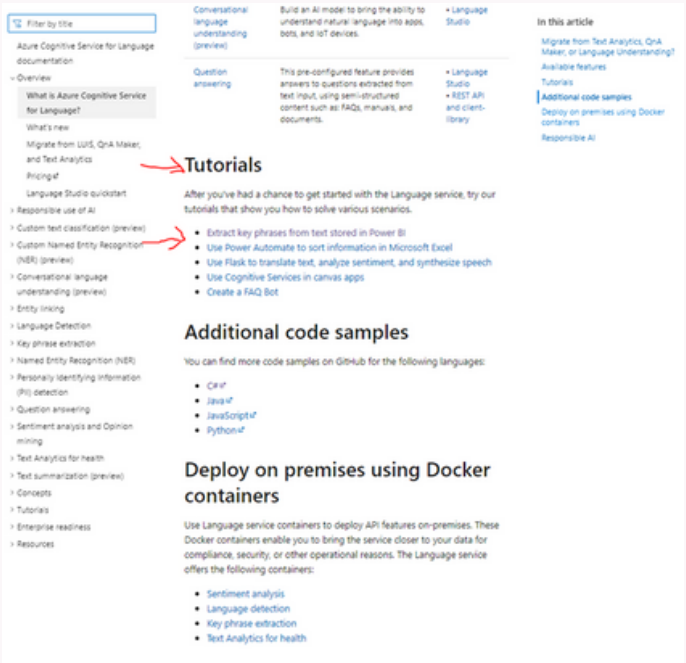
11



12



13



Techionista Holiday - Business Case

Create the Power Query function

Learn more about the dataset in this section of the guide.

Creating the Power Query function

Learn more about how to get started for the business case.

Create the Power Query function in the Query Editor

In Power BI, we will leverage a Power Query Function in order to run the rows of the dataset through the Cognitive Services in Azure.

As was mentioned on the previous page: make sure to leverage the Power Query Function on only 75 rows first! After you manage to get the function working, you can increase the number of rows in your dataset to the amount of data you'd like to use.

To set up the Power Query Function and run the Azure Cognitive Services, you need to follow additional steps. These steps can be found here.

Note: this link provides you with multiple options for custom functions. It's up to you which custom function(s) you'd like to create and use.

Why 75 rows?

Please read this carefully before invoking the custom function explained on the next page!

You have to be aware of the amount of rows you use in Power BI when connecting to Azure Cognitive Services. Azure charges money per row you call the Cognitive Services API upon. Your Azure Pass provides you with a **150$ budget**. You will be using the Standard tier of Cognitive Services, which costs **1.00$ per 1000 rows**.

The dataset you're using consists of about 515,000 rows! If you would call the API upon the complete set of rows you would blast through your 150$ budget immediately and then you can **not** invoke the custom function again.

Why 75 rows?

Since your budget is 150$, the maximum number of rows you can use to call the API upon, is **150,000 rows**. So make sure you first **keep 75 rows**, just to see if you can get everything to work properly when invoking the custom function. Then adjust the amount of rows to a dataset that you like to work with. So think about what rows of those 515,000 you'd like to use (ideas: pick certain countries, hotels, types of travelers, etc.). After you've chosen the dataset you want to work with, you can start invoking the custom function on that dataset.

Tip: you can check your Text Records usage in your Azure Portal under Monitoring > Metrics > Processed Text Records.

You can find more info on the pricing of Cognitive Services here.

Using 75 rows only will also save you a lot of time loading your data during the transforming and cleaning process. Once you are ready you can adjust the rows again to the amount you prefer. You can ease up this process by making use of a **parameter**. Check out this blog by Marc Lelijveld to see how it works.

Note: ignore the part of the blog about query folding, it is not applicable to this business case!

Final tip: first figure out what part of the dataset you want to use, THEN start building your report/dashboard, not the other way around!

Techionista Holiday - Business Case

The Build

Here you will learn more about what your final dashboard can look like and about your presentations.

Building your Dashboards

Learn more about how to get started for the business case.

Building

We won't tell you what the report/dashboard should look like, but we will give you some hints on what options may be important for the Techionista's Holidays managers.

Therefore you can follow the options below, but you're free to do as you please. You can modify the options below or add other visualizations if you think they add interesting insights. Remember to follow the corporate house style of Techionista's Holidays when building the report/dashboard.

- Ability to filter on different countries
- Ability to filter on different cities
- Ability to filter on different hotels
- Ability to filter on different dates
- Ability to filter on different type of trips
- Ability to filter on different types of travelers
- Ability to filter on different nationalities
- A map with an overview of all the hotel locations
- A visual that shows the division of i.e. different nationalities, traveler types, etc.
- Insights into the amount of positive and negative scores (tip: DAX SWITCH)
- The average sentiment score
- Word clouds with positive and negative key words
- The average sentiment score explained by i.e. city, type of trip, traveler type, etc. (tip: decomposition tree)
- Top 3 average lowest-scoring cities on the sentiment score/reviewer score
- A card visual with comments (tip: Strippet browser in the visual store, for this visual you need to make a simple index)

Besides making a report/dashboard you also start working with a detailed step-by-step plan and keep track of who does what, the choices you make and why.
You also write what feedback you have received during the presentation and what you have done with it.
In addition, you indicate how much time you spent on it, do certain subjects take more time than others?

Why do we ask you to make a detailed step-by-step plan? Because we like to know what steps you took, which choices you made and why, and of course how much time you spent on it. So please make note of this, because we would like to give you substantive feedback on how you made the report/dashboard.

Presenting your Business Case

Learn more about how to get started for the business case.

Presenting

You will have to present your results in a short pitch of 5 minutes on February **10th from 1 to 5 pm**. Your presentation will be approximately 10 minutes in total:

- the first 5 minutes you present your results: go over your visualizations, give a short demonstration of the report/dashboard, and explain your findings;
- In the second 5 minutes, you will answer questions from Techionista and receive feedback.

After all, groups have presented their results, there is a short break. After the break, we will give each group 10 minutes to receive questions and feedback from fellow students.

Your Audience

Your presentation should be aimed at end-users of the report/dashboard. They don't look at how it was made but look at how it looks and if it is functional.

Content-wise Techionista will give you feedback on how you made your report/dashboard at a later moment after you have submitted it for evaluation.

Good Luck and Have Fun!