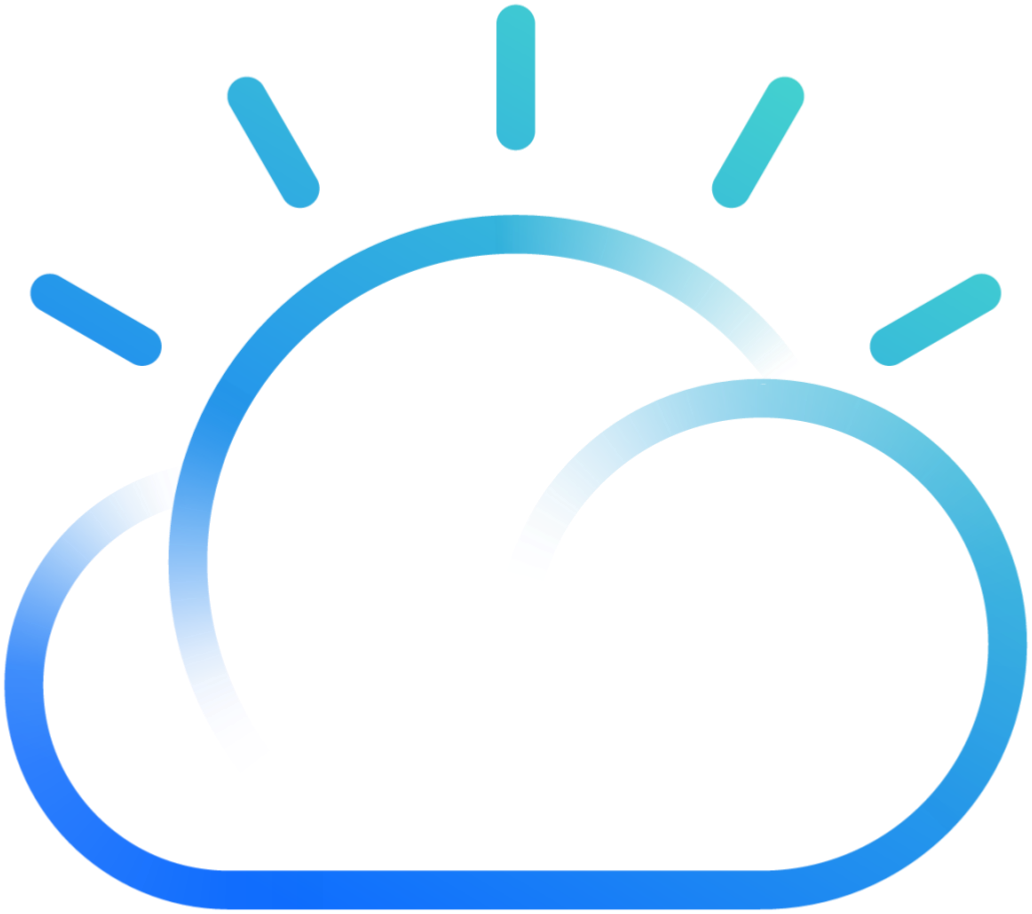
IBM Watson

# AI Lab Challenge 2

*Discovery Service: Unknown Knowns*



IBM

# Contents

# Preface

## Industry Use-case

According to Forbes, by 2020 about 1.7 MB of new information will be created every second—for each and every human being on the planet. This wealth of data offers extraordinary opportunities to:

- Provide your organization with insights extracted from exploding volumes of data.
- Integrate growing volumes of unstructured data to reveal previously undiscovered trends.
- Rapidly scale the learning curve and overcome three obstacles organizations face when deploying cognitive solutions: identify and ingest data sources; enrich data so it's searchable; and build queries.

In this lab, you will gain insights from actual AirBnB reviews of Manhattan apartments. A typical query from Yelp or other review engines may depict a certain apartment as most desirable to stay at, however, for example, the reviews may not mention that there is a bar nearby and the noise level during closing hours are enough to wake up any of the residents in the vicinity apartments; the Discovery service can surface those concerns.

Other instances may be to collect reviews with a negative sentiment and perhaps as AirBnB administrators to inform the tenant of the less-than stellar reviews that they may have been receiving and perhaps have not taken any action to better the situation for the renters.

## Overview

IBM Watson™ Discovery makes it possible to rapidly build cognitive, cloud-based exploration applications that unlock actionable insights hidden in unstructured data — including your own proprietary data, as well as public and third-party data.

This is the architecture of a complete Discovery service solution:


information you need, and then integrate those insights into your new application or existing solution.

How does Discovery do it? By using data analysis combined with cognitive intuition to take your unstructured data and enrich it so you can discover the information you need.

**Estimated Time to Complete:** 60 mins

## Prerequisites

This lab requires that you already have an IBM Cloud Account in order to utilize the Watson services.

## Objectives

There are four Milestones you must complete:

1. Create Discovery Service
2. Create Collection
3. Upload and Enrich Data
4. Explore Negative Sentiments

## Tools

IBM Cloud

Watson Discovery Service

IBM's **Watson Discovery Service** is a suite of APIs that aims to make it easier for companies to ingest and analyze their data, even if they don't have an advanced degree in data science.

## Watch the Video

The following image is a link to a video explaining the steps involved with your discovery Journey.

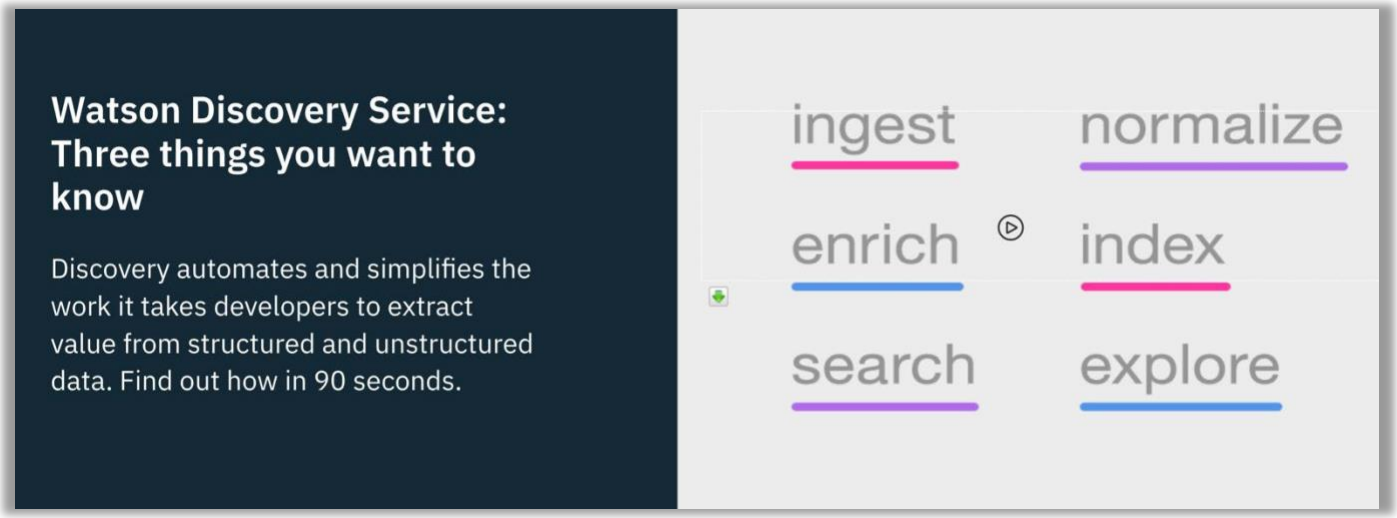



*Figure 0-2* **Watson Discovery Service – Video Link**

It is strongly recommended that you watch this video, https://www.youtube.com/watch?v=q4SVFg6bN5E

# Milestone 1: Create Discovery Service

## Milestone Overview

This lab requires you to complete four Milestones:

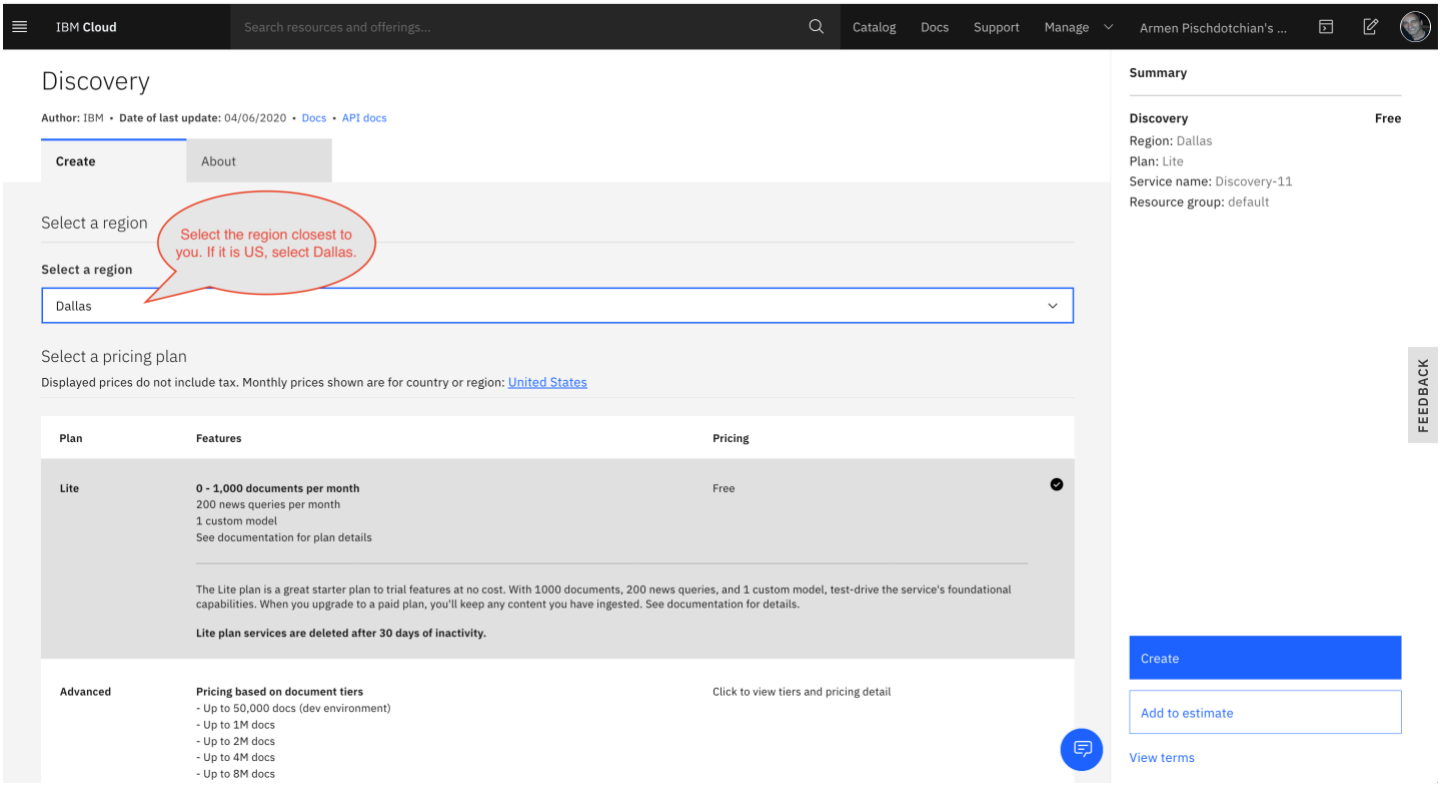
1. **Create Discovery Service**
2. Create Collection
3. Upload and Enrich Data
4. Explore Negative Sentiments

In this Milestone we will navigate to the Cloud Catalog and create our discovery service.

## Steps

1. Login into IBM Cloud: https://cloud.ibm.com
2. Click the **Catalog** tab.
3. Search for the **Discovery** service and click that tile.

4. Edit the Service name to something meaningful to you (for example: Discovery-mydiscovery) and select a region closest to where you are and click **Create** (If you have just created your account and accessed it from the confirmation email, you may need to log into IBM Cloud once again, then you can see the Create button in the bottom right corner).




5. Click the **Launch Watson Discovery** button.

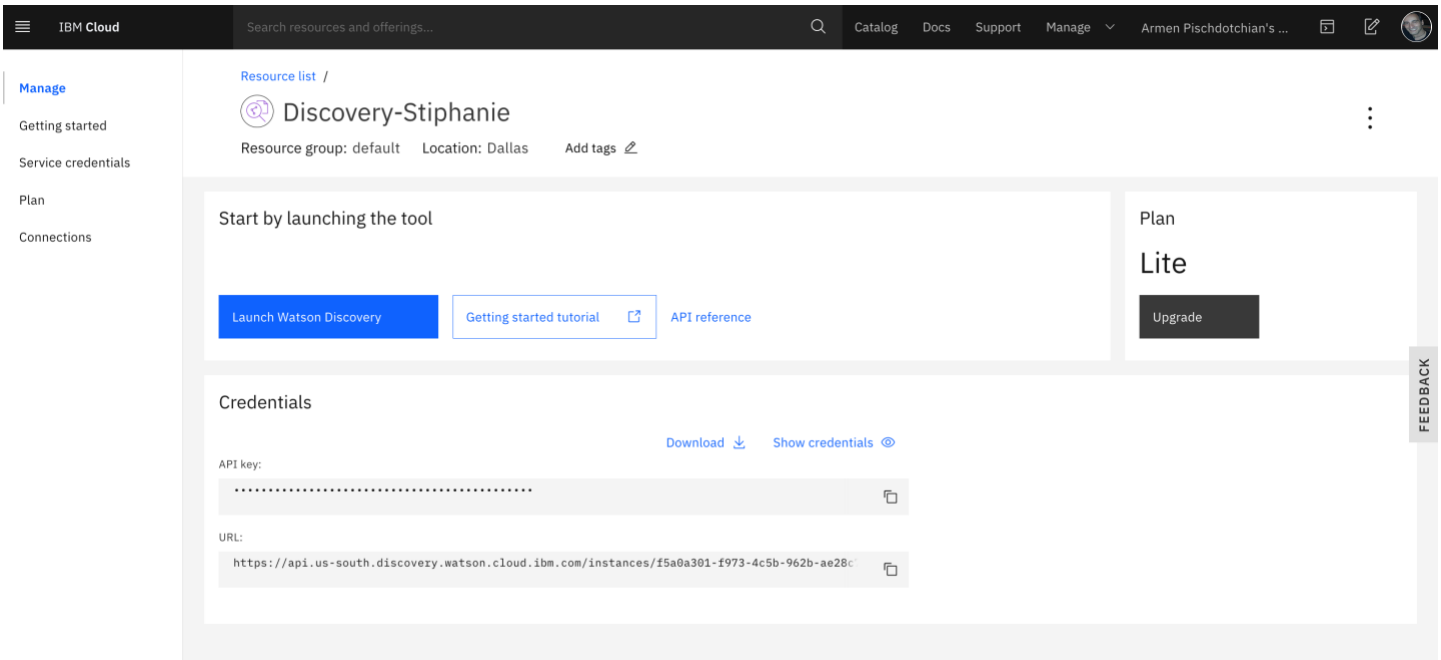



***Figure 1-2*** **Discovery Dashboard**

6. Click **Let's get started**.
7. Click **Upload your own data**.
8. Click **Continue**, twice.

## Milestone Summary

You have now created your Discovery service and are ready to begin creating Collections that will process our content.

# Milestone 2: Create Collection

## Milestone Overview

This lab requires you to complete four Milestones:

1. Create Discovery Service
2. **Create Collection**
3. Upload and Enrich Data
4. Explore Negative Sentiments

Before you add your own content to the Discovery service; the best practice is to configure the service to process the content the way that you want. We will create an AirBnB Collection.

## Steps

1. Click **Upload your own data**.
2. Click **Set up with current plan**.
3. Click **Continue**.

4. Name your collection; for example: AirBnB

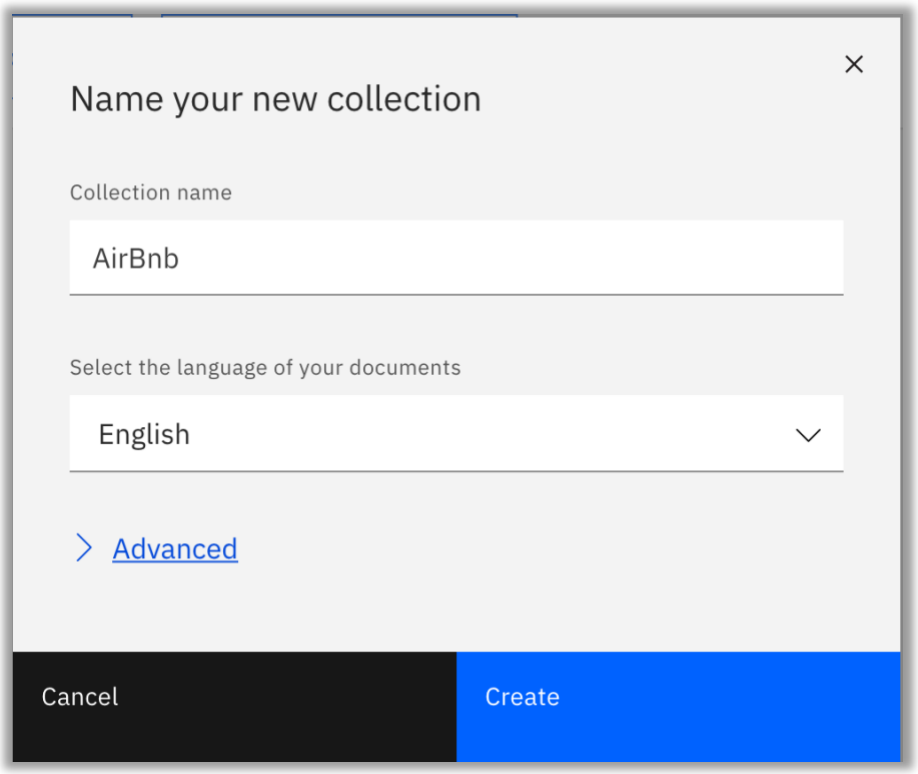



*Figure 2-1* **Creating an AirBnB Collection**

5. Click **Create**.

## Milestone Summary

We have now created our Collection and it is time to begin adding data.

# Milestone 3: Upload and Enrich Data

## Milestone Overview

This lab requires you to complete four Milestones:

1. Create Discovery Service
2. Create Collection
3. **Upload and Enrich Data**
4. Explore Negative Sentiments

Each collection you create is a logical division of your data in the environment. Each collection will be queried independently when you get to the point of delivering results. Why would I want more than one collection? There are a few reasons, including:

- You may want multiple collections in order to separate results for different audiences.
- The data may be so different that it doesn't make sense for it all to be queried at once.

## Steps

1. Download and then extract the AirBnB reviews from this repository, https://github.com/apischdo/Artificial-Intelligence-and-Data-Science/blob/master/AirBnB_rental_reviews.zip
2. Click **Upload your data** and browse to where you downloaded the artifacts of the zip file. There are 101 AirBnB reviews of Manhattan apartments.

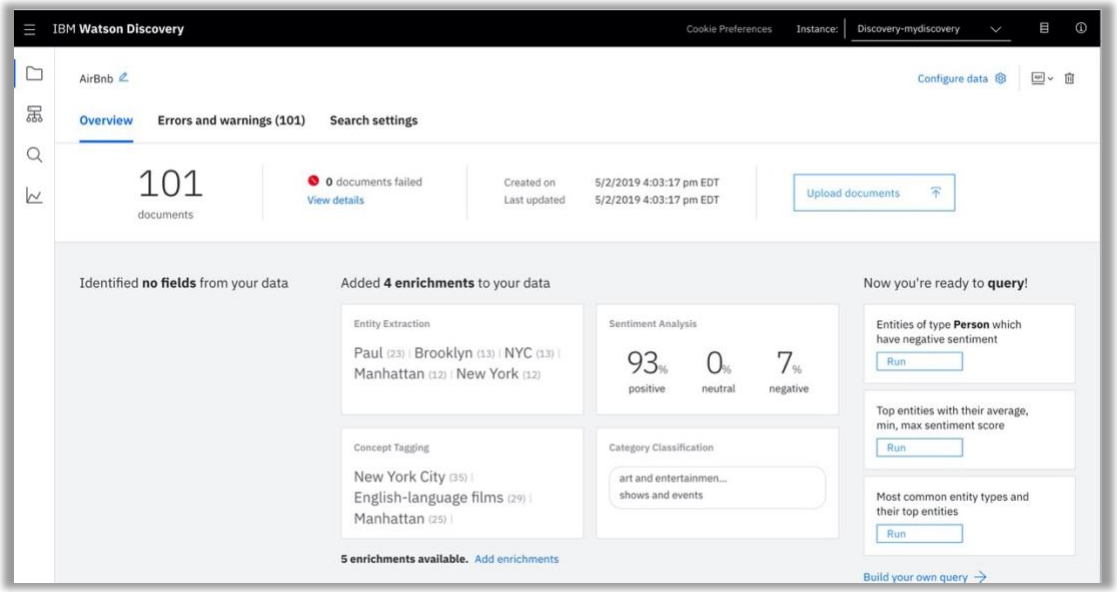



*Figure 3-1* **Watson Discovery Overview**

3. Click **Add enrichments**.
4. Click **Add enrichments** again in the ensuing page.
5. Add the following three enrichments:
    a. Keyword Extraction
    b. Emotion Analysis
    c. Relation Extraction

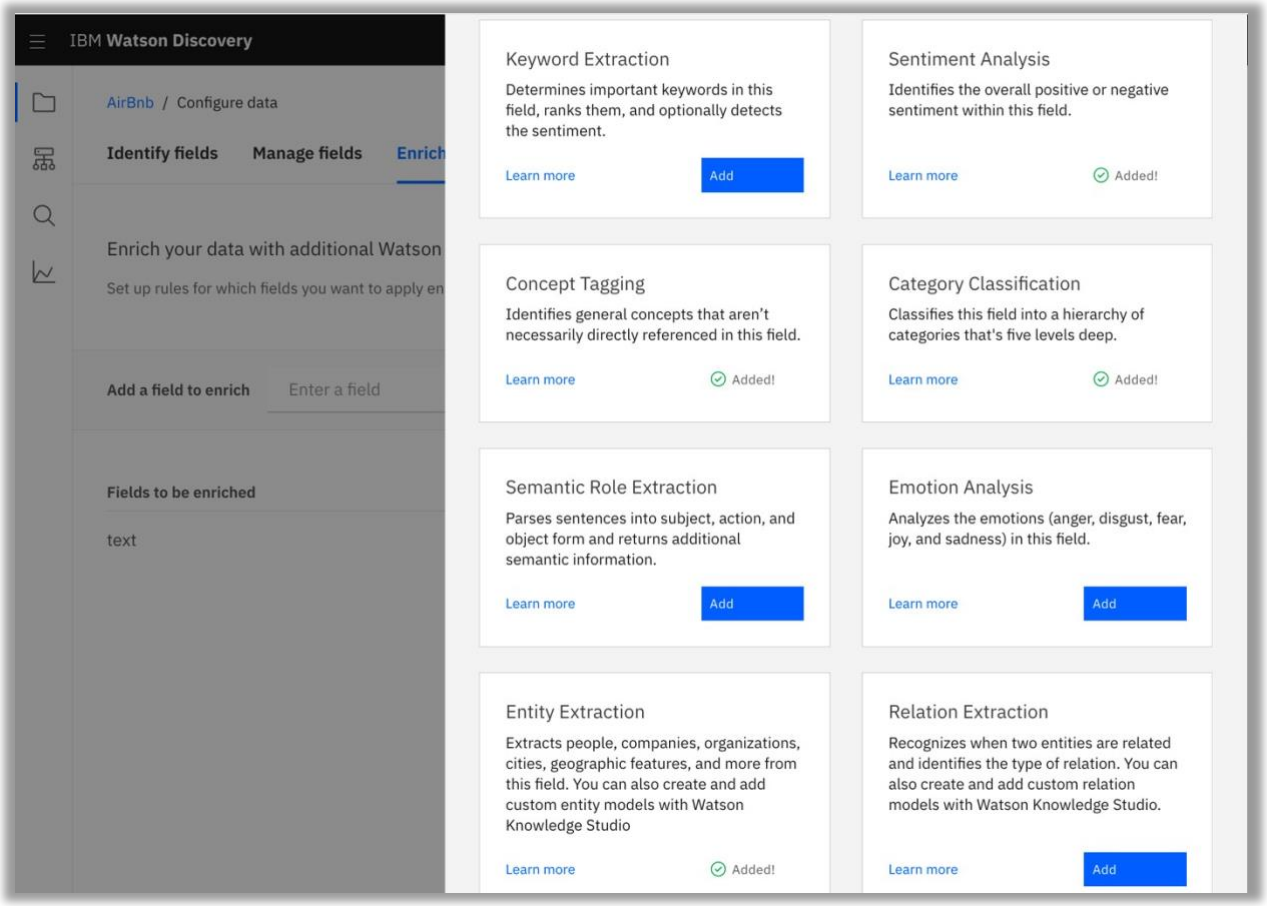



***Figure 3-2*** **Add enrichments**

6. Close the dialog box.
7. Click **Apply changes to collection**, but do not upload any new content.

8. Click the **folder Icon** in the left panel and click the your AirBnB collection.

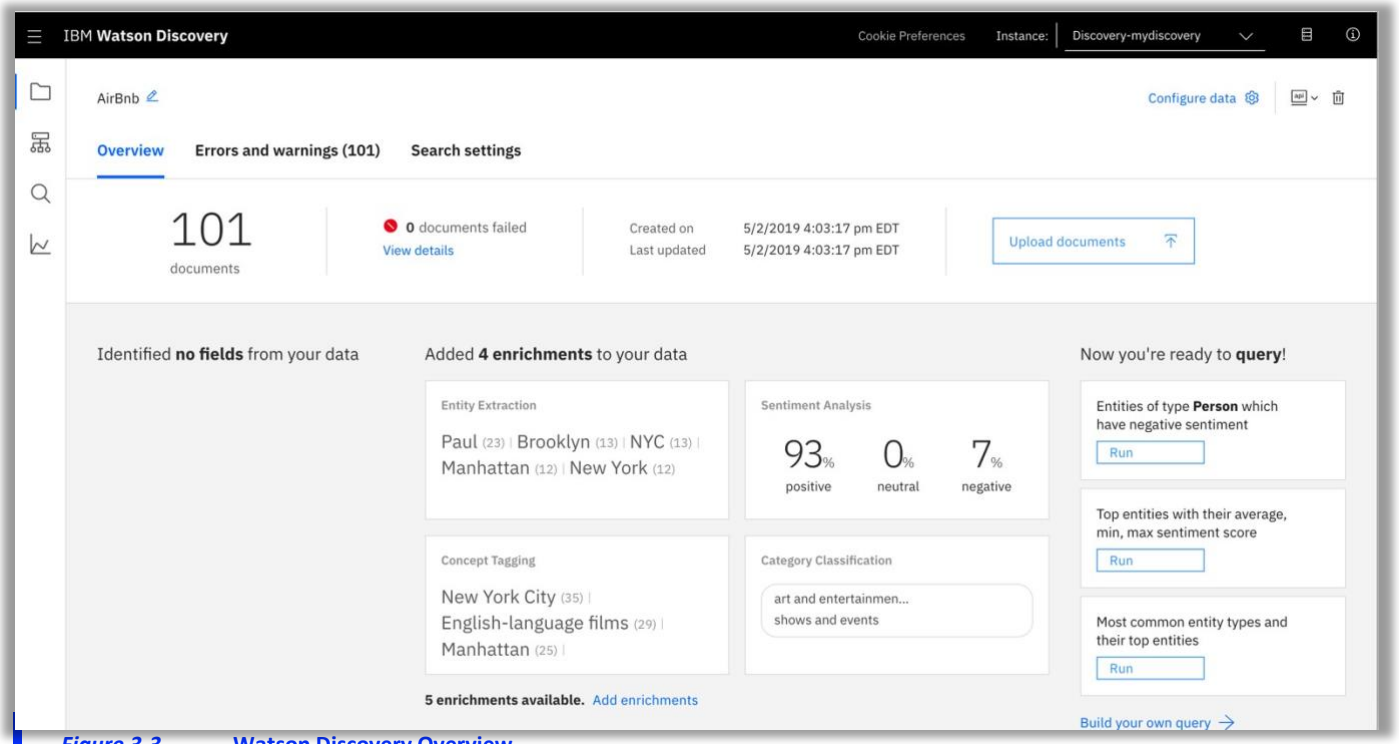



*Figure 3-3* **Watson Discovery Overview**

## Milestone Summary

Completion of this Milestone means that data has been successfully added to our AirBnB Collection. In the next Milestone, we will dive into the data to find user sentiment.

# Milestone 4: Explore Negative Sentiments

## Milestone Overview

This lab requires you to complete four Milestones:

1. Create Discovery Service
2. Create Collection
3. Upload and Enrich Data
4. **Explore Negative Sentiments**

Notice that you have 7 negative sentiments regarding the AirBnB reviews.

## Steps

1. Click the **Schema Icon** in the left panel.
2. Select **Sentiment enrichment**.

Notice the hierarchy of the enriched_text field:

**enriched_text.sentiment.document.label**

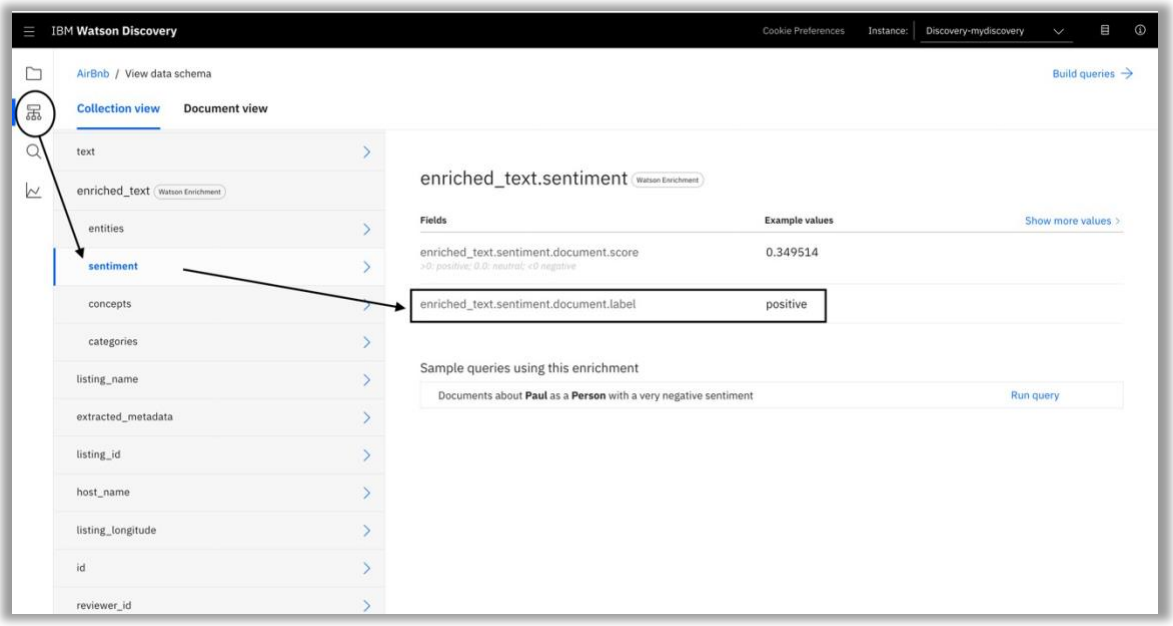



*Figure 4-1* **Watson Discovery – Sentiment Enrichment**

You are now going to use the same path in your query, except that you specify a negative value.

3. Click the **Query Icon** in the left panel.
4. Run the Query without specifying any parameters just to ensure you can see all 101 artifacts.
5. Open the **Search for documents field** and type in the nature language section**: cold room** and run the Query.
6. Click the **JSON tab**.

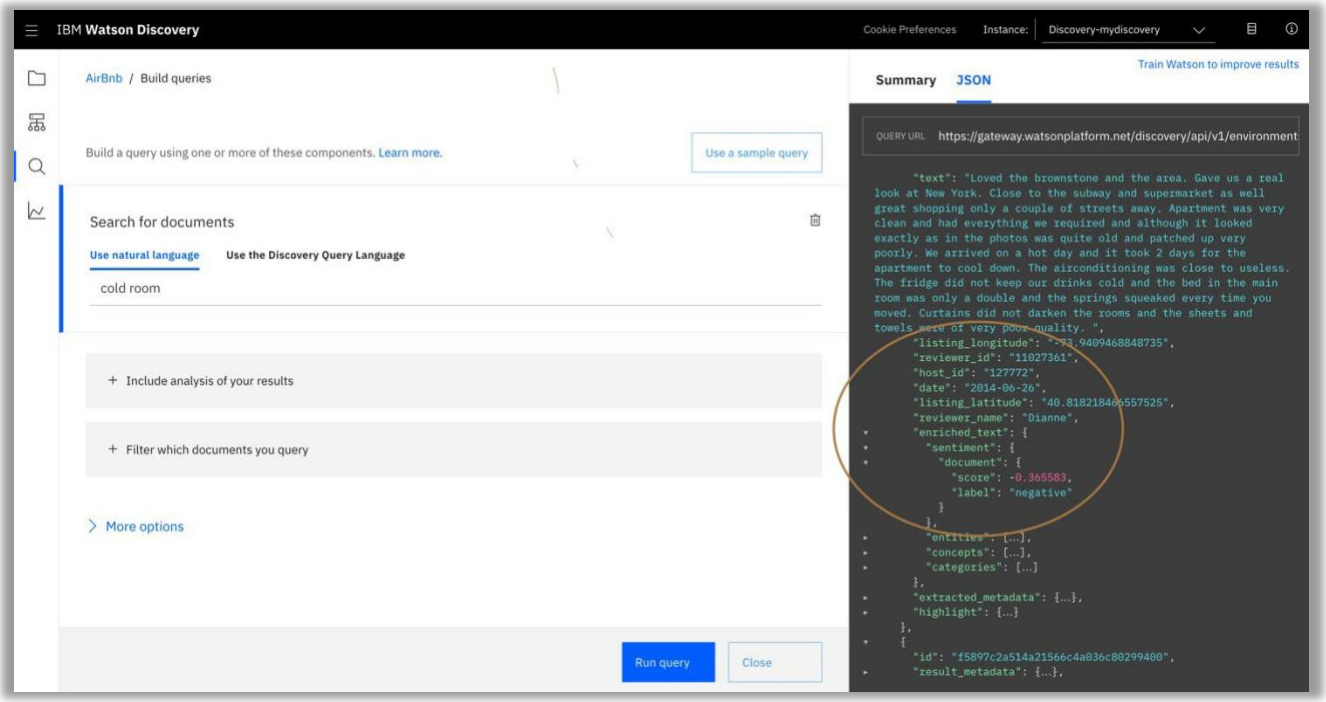



*Figure 4-2* **Watson Discovery JSON**

7. Scroll down until you see the **enriched_text** field and expand that, more specifically the **sentiment sub-field**.

However, your query had 50 results and that does not exactly zero in on the 7 negative sentiment types that you saw in the Schema, so let's use the query language.

8. Click the **Trash Icon** next to the natural language field (cold room) and remove that to see all 101 reviews.
9. Open the **Filter** which documents the query section. Use the drop down to specify:

**enriched_text.sentiment.document.label is negative.**

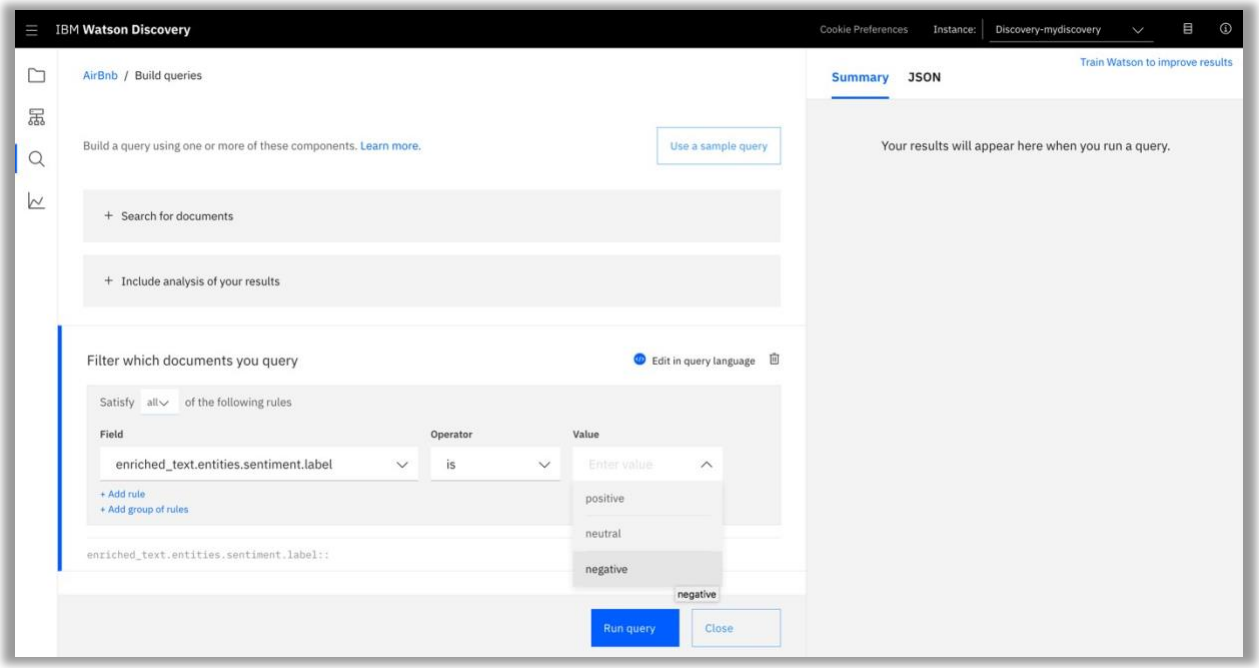



***Figure 4-3*** **Build Queries**

10. Click **Run Query**.

11. Open the enriched_text filed and note the negative label and the confidence value(score) associated with the negative label.

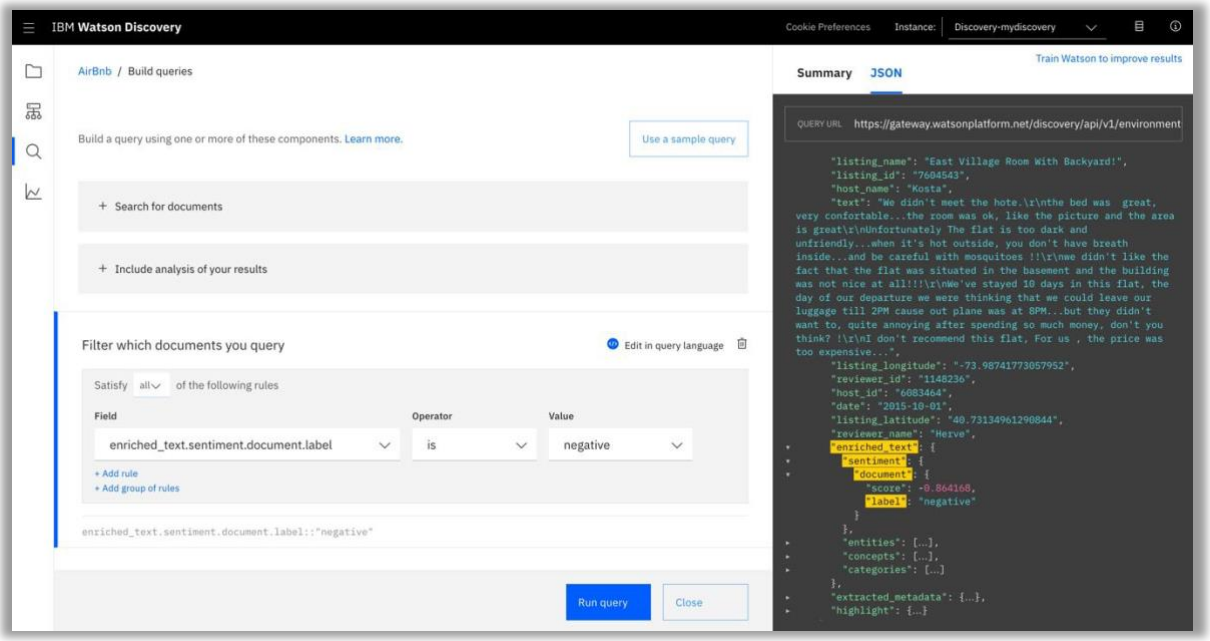



***Figure 4-4*** **Note the Enriched Text**

## Milestone Summary

In this Milestone, we explored our data and pinpointed Negative Sentiment. We then learned how to run further queries and how to provide Watson with further training. Congratulations, you have now completed *Gaining Insights from AirBnB*!