

School of Information Studies  
Syracuse University  
**IST 615 – Cloud Management**  
**Project Progress Report**  
**Car Retail Website**



**Due Date:** 11/14/2022

**Submission Date:** 11/14/2022

**Number of Pages:** 06

**Team members:**

- Mikhail Pravin Pinto ([mpinto01@syr.edu](mailto:mpinto01@syr.edu))
- Kirat Saran ([ksaran@syr.edu](mailto:ksaran@syr.edu))

**Contents**

<b>S.No.</b>	<b>Description</b>	<b>Page No.</b>
1.	Introduction	2
2.	Progress	2
3.	Future Implementations	6

## Introduction:

We are building a car retail website using cloud where buyers can buy cars based on manufacturers, mileage, distance travelled, etc since the demand to buy vehicles has increased post pandemic.

The dataset is from the site CarDekho.com which was uploaded to Kaggle.com.

Link to the dataset:

<https://www.kaggle.com/datasets/austinreese/craigslist-carstrucks-data?datasetId=62920>

We'll use Microsoft Azure Cloud Services to implement this.

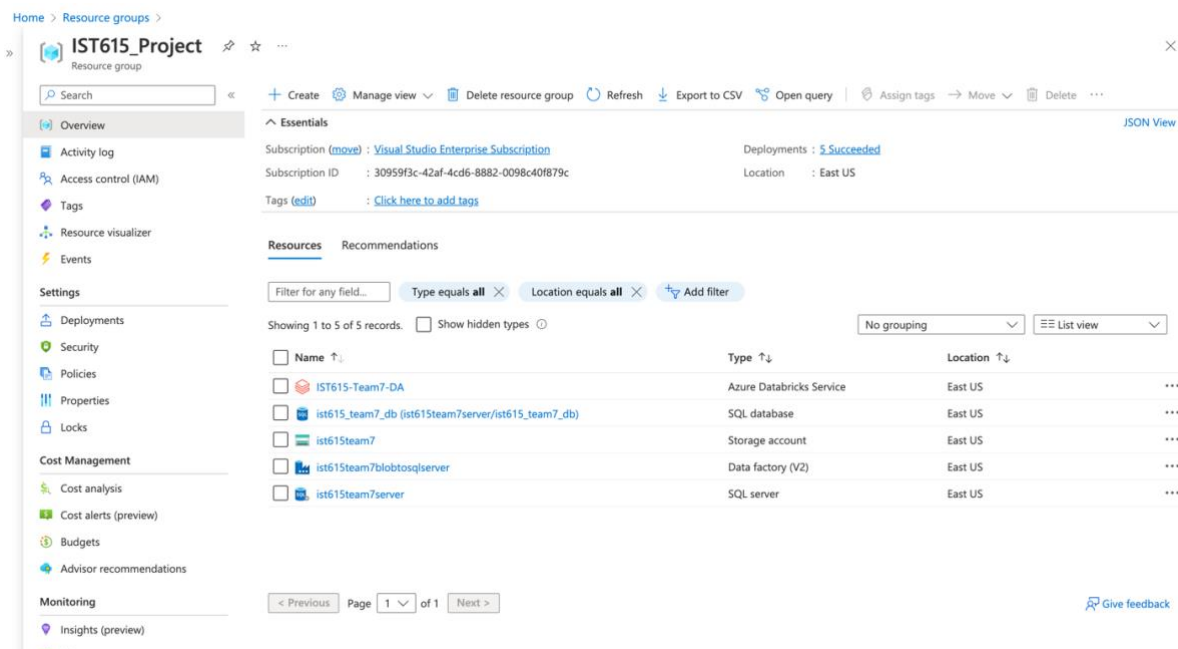
## Progress:

### *1. Data Cleaning*

CarDekho.com dataset was downloaded in the form of a csv file. We cleaned the data by removing rows with null values in columns like condition, manufacturer, model, fuel, odometer, and transmission.

Column with County name was also dropped as it had more than 50% null values. Similarly, the size and region URL was also removed due to a lot of missing values. The description was also dropped as the column contained identical descriptions for more than 50% of the records.

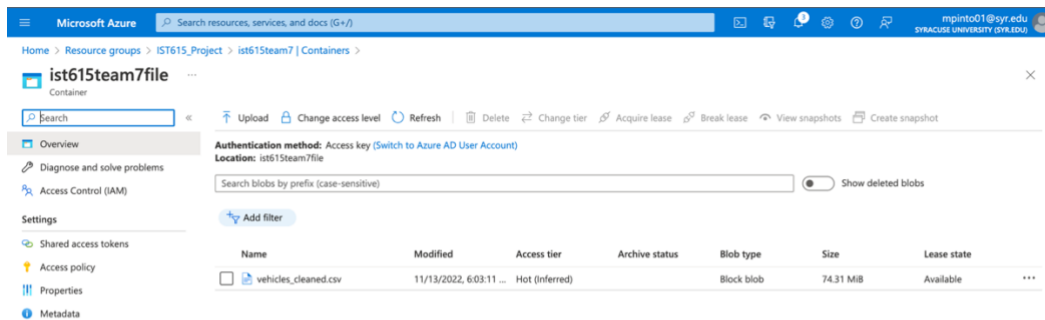
We created a resource group for the project to manage and associate all the multiple resources that would be used for this project.



*Screenshot 1: Resource group created for the project.*

## 2. Azure Blob Storage

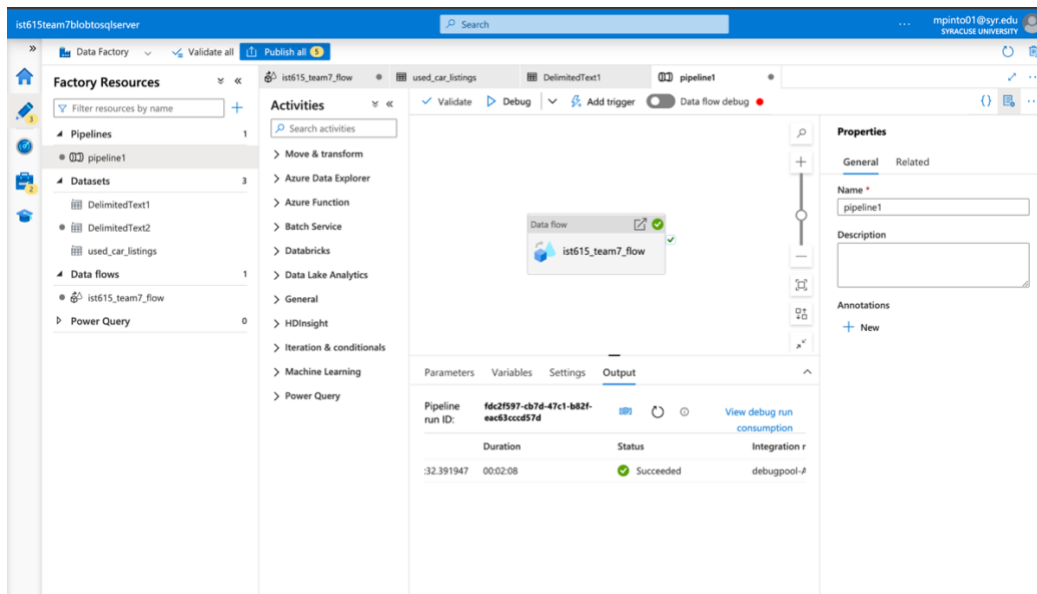
Our cleaned dataset file 'vehicles\_cleaned.csv' was uploaded to Azure blob service and stored in a container.



Screenshot 2: A container created in Azure blob storage to upload and store our cleaned csv data file.

## 3. Azure Data Factory

Azure Data Factory has been used to facilitate data integration and data transformation. A data flow was created to get data from the Azure Blob storage container to Azure SQL Server DB. This was executed using a pipeline to get data from Azure blob storage to SQL Server.



Screenshot 3: A pipeline created and executed in Azure data factory to get data from Azure blob storage to SQL Server.

**ist615\_team7\_db (ist615team7server/ist615\_team7\_db)**

SQL database

Search

Copy Restore Export Set server firewall Delete Connect with... Feedback

**Overview**

Activity log

Tags

Diagnose and solve problems

Getting started

Query editor (preview)

**Settings**

Compute + storage

Connection strings

Properties

Locks

**Data management**

Replicas

Sync to other databases

**Integrations**

Stream analytics (preview)

Add Azure Search

**Power Platform**

Power BI

**Essentials**

Resource group (move): IST615\_Project

Status: Paused

Location: East US

Subscription (move): Visual Studio Enterprise Subscription

Subscription ID: 30959f3c-42af-4cd6-8882-0098c40f879c

Tags (edit): Click here to add tags

Server name: ist615team7server.database.windows.net

Connection strings: Show database connection strings

Pricing tier: General Purpose - Serverless: Gen5, 1 vCore

Auto-pause delay: 1 hour

Earliest restore point: 2022-11-13 22:13 UTC

**Getting started** Monitoring Properties Features Notifications (1) Integrations Tutorials

**Start working with your database**

Connect to your database and start working with data with a few simple steps. [Learn more](#)

**Configure access**  
Configure network access to your SQL server. [Learn more](#)

**Connect to application**  
Use connection strings to connect to your SQL database from your applications and favorite tools.

**Start developing**  
Work in your database by using tools to add, modify and query data. [Compare tools](#)

[Configure](#) [See connection strings](#) [Open Azure Data Studio](#)

Open in Visual Studio

Screenshot 4: Azure SQL server and DB created for the purpose of data handling and data transformation.

**ist615\_team7\_db (ist615team7server/ist615\_team7\_db) | Query editor (preview)**

SQL database

Search

Login + New Query Open query Feedback

Query 1 Query 2

Run Cancel query Save query Export data as Show only Editor

```
1 SELECT TOP (1000) * FROM [dbo].[usedcar]
```

**Results** Messages

Search to filter items...

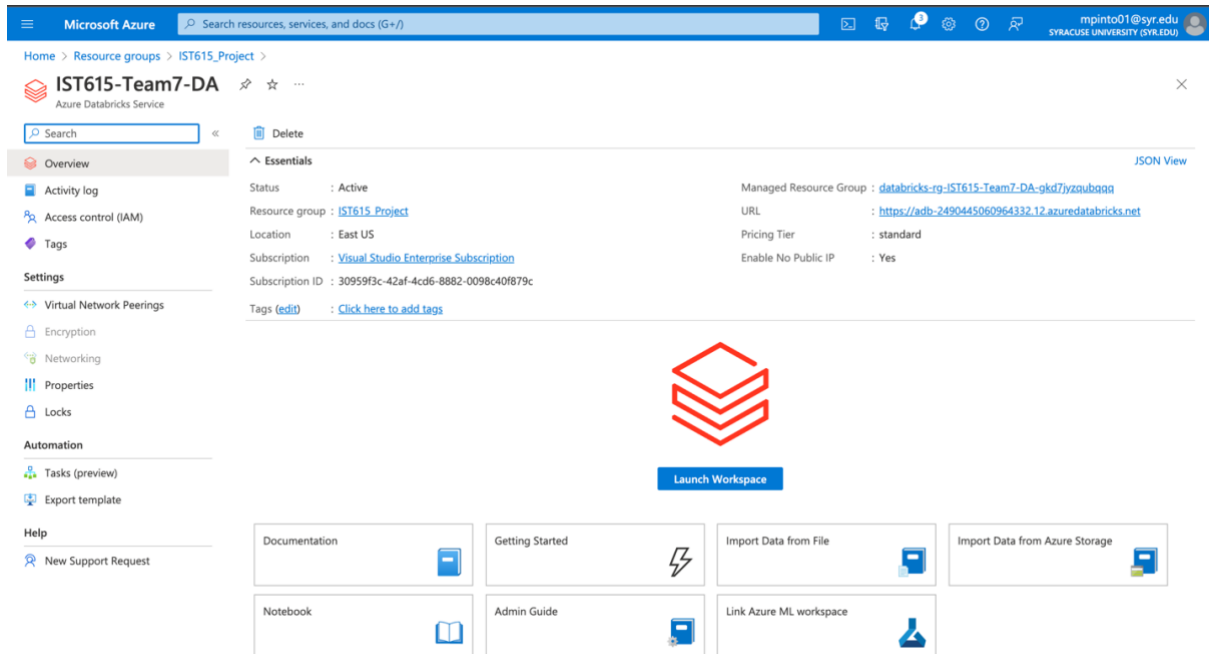
id	url	region	price	year	manufacturer
7307193415	https://sd.craigslist.org/cto/d/...	south dakota	17900	2019.0	chevrolet
7306278938	https://sd.craigslist.org/ctd/d/s...	south dakota	28500	1989.0	rover
7305950913	https://sd.craigslist.org/ctd/d/r...	south dakota	7995	2000.0	ram
7305940960	https://sd.craigslist.org/cto/d/s...	south dakota	34500	1969.0	chevrolet
7305437418	https://sd.craigslist.org/ctd/d/f...	south dakota	47950	2020.0	ford
7305437144	https://sd.craigslist.org/ctd/d/f...	south dakota	65950	2020.0	ford

Query succeeded | 0s

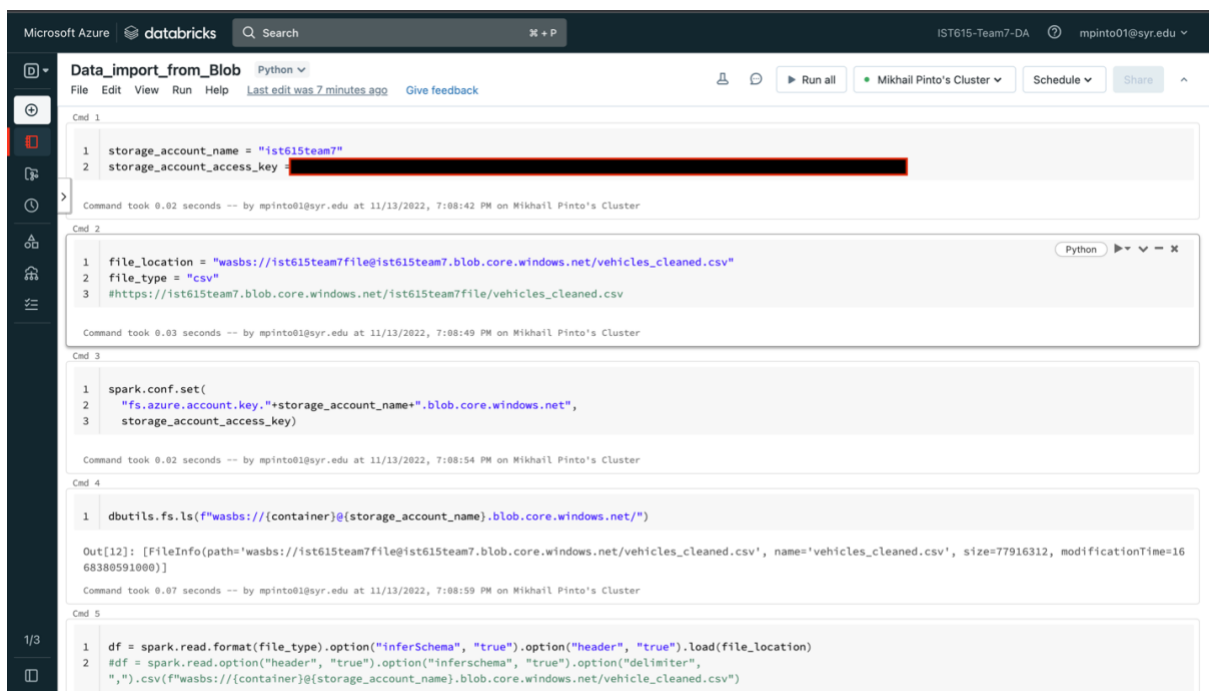
Screenshot 5: An SQL query to display top 1000 rows out of 234k total records.

#### 4. Azure Databricks

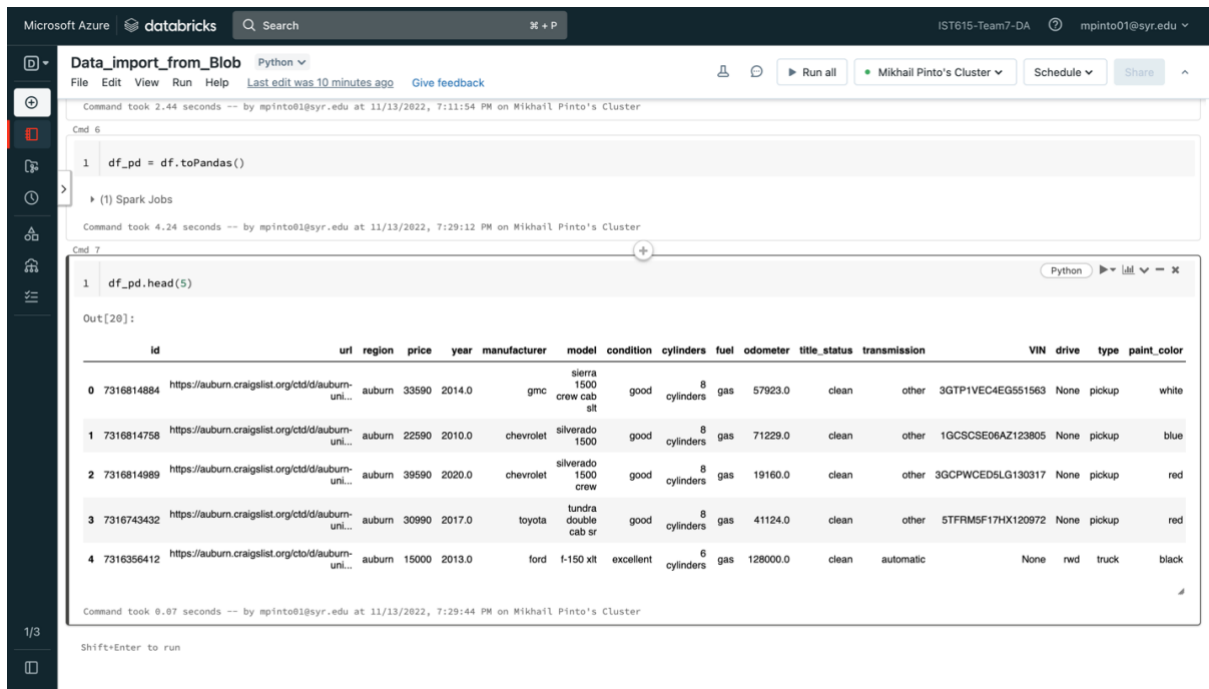
This would be used for the purpose of python scripting, analysis and model creation in our project. We wrote a script to integrate Azure Blob storage and Azure Databricks. Using this script, we can read data from our container in blob storage to do further analysis on our dataset.



Screenshot 6: Setting up Azure Databricks service for our project.



Screenshot 7: Script written to integrate Azure Blob storage and Azure Databricks



Screenshot 7: Converting our data to a pandas dataframe using databricks service.

## Future Implementations:

S.No.	Implementation	Expected Completion
1.	<p>Implemenation of basic front end</p> <ul style="list-style-type: none"> <li>Search Page</li> </ul> <p>This page will have user inputs like the brand of the car, the transmission of the car, the type of car, the color of the car, price of the car. This will retrieve car data for that brand as well as cars with similar characteristics</p> <ul style="list-style-type: none"> <li>Post Page</li> </ul> <p>According to the price of the cars already present in the dataset, we will estimate the price of user's car if a user wants to know how much their car is valued in the market. If the user likes the estimated price, they can list their car on our website. This would add the data to the database in the SQL server.</p>	November 21st – 30 <sup>th</sup> , 2022.
2.	Create PowerBI Dashboards based on our analysis and modelling	December 1 <sup>st</sup> , 2022.
3.	Ready with our final project and presentation	December 5 <sup>th</sup> , 2022.