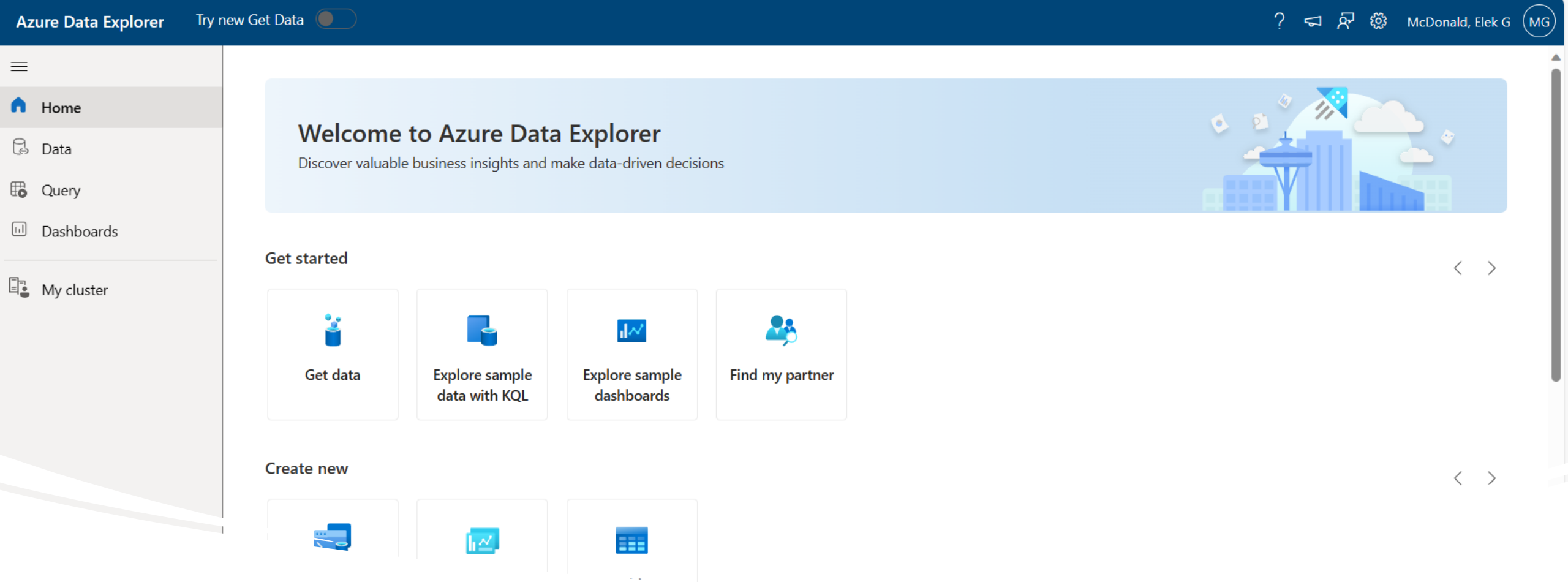
Azure Final Project: Healthcare Data

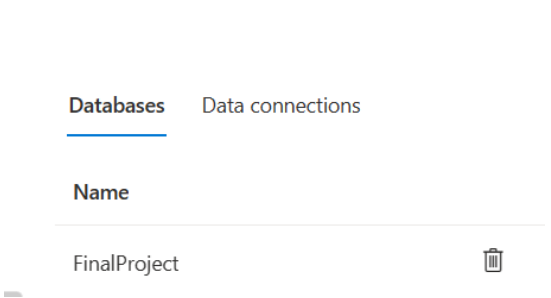
Cole, Larry, Elek

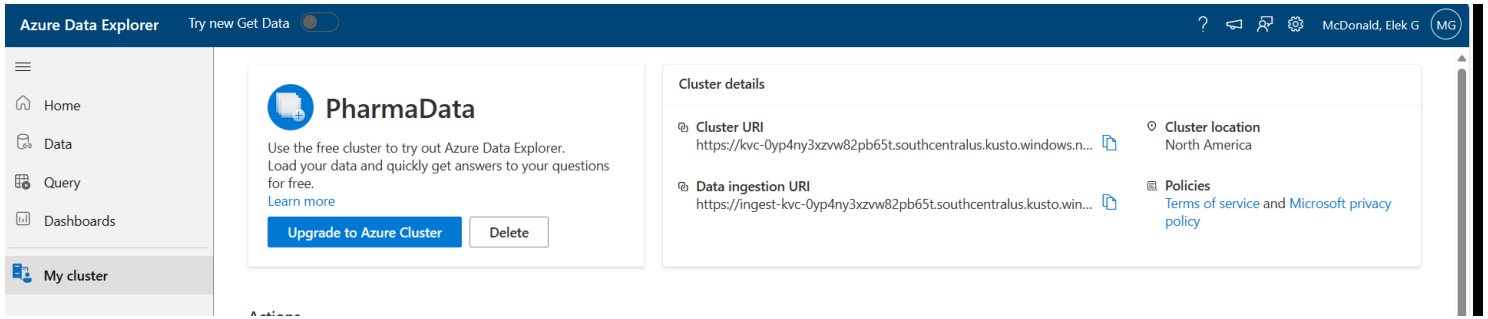


Azure Data Explorer

- Settled on Azure Data Explorer as our service
- Goal = load data, run queries, make visuals

Made Cluster and Database

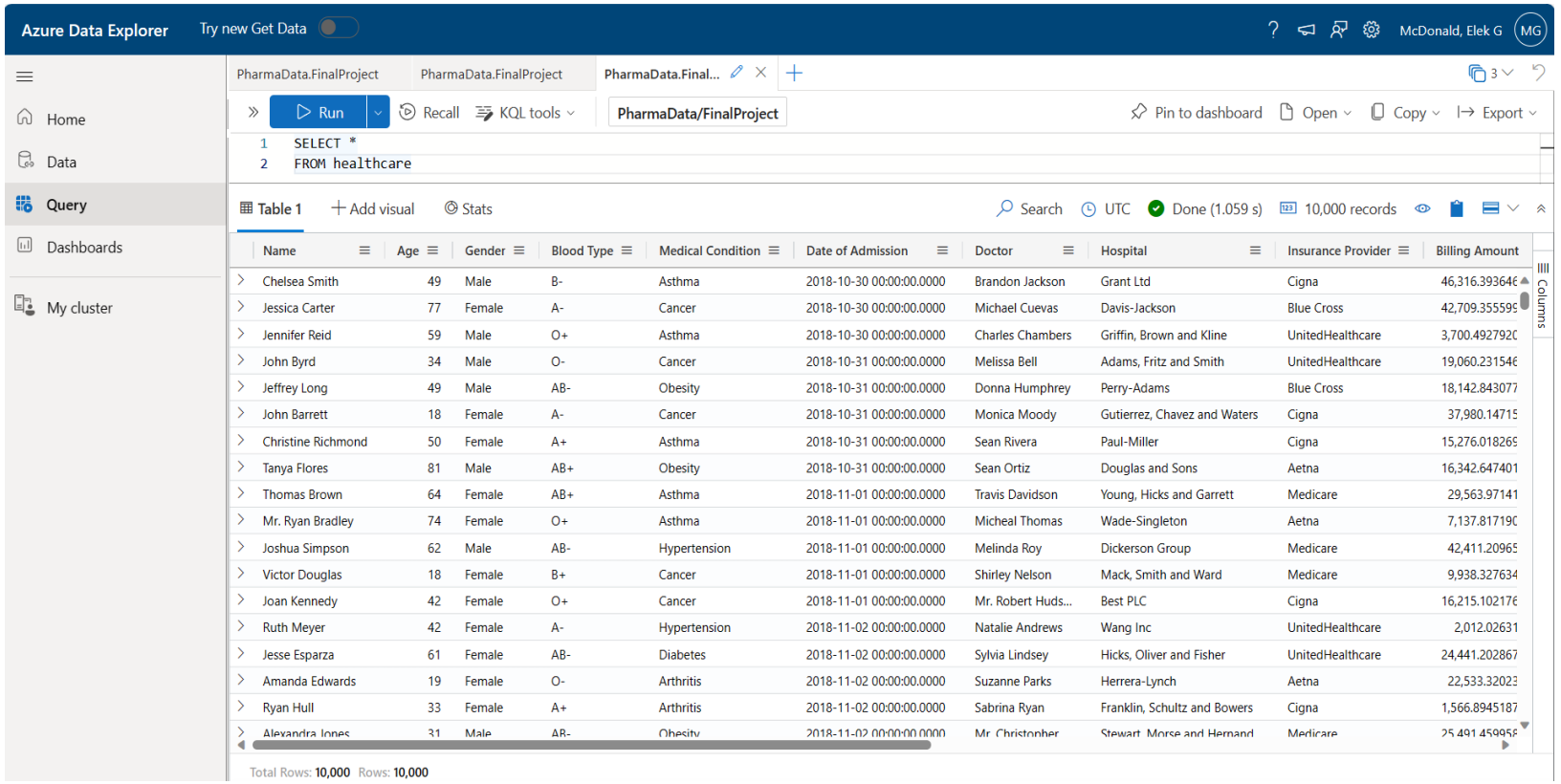




- To begin the query → made a cluster
  - Made database to upload csv files as needed
- Started with pharma dataset
  - Switched to healthcare dataset for visualization purposes
    - Pharma data was confusing → numerical data didn't make sense

Ran Query

- Ran a SELECT statement as our initial query to visualize our dataset



Name	Age	Gender	Blood Type	Medical Condition	Date of Admission	Doctor	Hospital	Insurance Provider	Billing Amount
Chelsea Smith	49	Male	B-	Asthma	2018-10-30 00:00:00.0000	Brandon Jackson	Grant Ltd	Cigna	46,316.39364€
Jessica Carter	77	Female	A-	Cancer	2018-10-30 00:00:00.0000	Michael Cuevas	Davis-Jackson	Blue Cross	42,709.355595
Jennifer Reid	59	Male	O+	Asthma	2018-10-30 00:00:00.0000	Charles Chambers	Griffin, Brown and Kline	UnitedHealthcare	3,700.492792C
John Byrd	34	Male	O-	Cancer	2018-10-31 00:00:00.0000	Melissa Bell	Adams, Fritz and Smith	UnitedHealthcare	19,060.23154€
Jeffrey Long	49	Male	AB-	Obesity	2018-10-31 00:00:00.0000	Donna Humphrey	Perry-Adams	Blue Cross	18,142.843077
John Barrett	18	Female	A-	Cancer	2018-10-31 00:00:00.0000	Monica Moody	Gutierrez, Chavez and Waters	Cigna	37,980.14715
Christine Richmond	50	Female	A+	Asthma	2018-10-31 00:00:00.0000	Sean Rivera	Paul-Miller	Cigna	15,276.018269
Tanya Flores	81	Male	AB+	Obesity	2018-10-31 00:00:00.0000	Sean Ortiz	Douglas and Sons	Aetna	16,342.647401
Thomas Brown	64	Female	AB+	Asthma	2018-11-01 00:00:00.0000	Travis Davidson	Young, Hicks and Garrett	Medicare	29,563.97141
Mr. Ryan Bradley	74	Female	O+	Asthma	2018-11-01 00:00:00.0000	Micheal Thomas	Wade-Singleton	Aetna	7,137.81719C
Joshua Simpson	62	Male	AB-	Hypertension	2018-11-01 00:00:00.0000	Melinda Roy	Dickerson Group	Medicare	42,411.20965
Victor Douglas	18	Female	B+	Cancer	2018-11-01 00:00:00.0000	Shirley Nelson	Mack, Smith and Ward	Medicare	9,938.327634
Joan Kennedy	42	Female	O+	Cancer	2018-11-01 00:00:00.0000	Mr. Robert Huds...	Best PLC	Cigna	16,215.102176
Ruth Meyer	42	Female	A-	Hypertension	2018-11-02 00:00:00.0000	Natalie Andrews	Wang Inc	UnitedHealthcare	2,012.02631
Jesse Esparza	61	Female	AB-	Diabetes	2018-11-02 00:00:00.0000	Sylvia Lindsey	Hicks, Oliver and Fisher	UnitedHealthcare	24,441.202867
Amanda Edwards	19	Female	O-	Arthritis	2018-11-02 00:00:00.0000	Suzanne Parks	Herrera-Lynch	Aetna	22,533.32023
Ryan Hull	33	Female	A+	Arthritis	2018-11-02 00:00:00.0000	Sabrina Ryan	Franklin, Schultz and Bowers	Cigna	1,566.8945187
Alexandra Jones	31	Male	AB-	Obesity	2018-11-02 00:00:00.0000	Mr. Christopher	Stewart, Morse and Hernand	Medicare	25,491.459958

Visualization Attempt

- Had some experience in SQL, but no training in KQL, so we had to learn syntax through ChatGPT, Bing, and google

- Attempted first visual
- Confusing graph about Billing Amount and Age

Overview – Our Challenges with KQL Coding

- **After ingesting our data into Azure Data Explorer, problems began to arise once we attempted to create visualizations. The obstacles goes as follows:**
- **Complex Queries:**
  - KQL is a powerful query language, but constructing complex queries for visualizations can be challenging. Ensuring that the queries meet the specific requirements for visualization might require a deep understanding of both the data and the KQL syntax. That being said, the KQL syntax posed a large challenge to us as all of the different queries we tried to utilized came out as "errors"
- **Data Formatting Issues:**
  - Visualizations often require data to be formatted in a specific way. After ingesting our csv file into our Azure cluster, the data within our csv contained specific flaws (spaces in column names) which required us to reformat our csv file so that we may use the data within Data Explorer.

To address these challenges, it is important that we utilize our resources (internet, other classmates who have knowledge within the language) to understand the intricacies of KQL and explore the capabilities of any associated visualization tools or platforms. Additionally, seeking assistance from online communities or documentation resources can be valuable in overcoming specific hurdles.