Exploring Microsoft Azure

By: Ben Phillippy, Matt Chylack, Shane Artis

Project and Presentation Outline

Step 1: Completed AZURE Trainings

Step 2: Identified our Dataset and Key Azure Trainings to use for Project

Step 3: Discovered and Uploaded Data to Microsoft Data Explorer

Step 4: Created Visualizations to better understand dataset and trends

Overview

Step 1: Completed AZURE Trainings

AZURE Trainings

Between the three of us completed trainings within; Understanding Data Concepts, Designing effective Power BI reports, Fundamentals of Microsoft Dynamic 365 Supply Chain Management, **Data visualization with Azure Data Explorer**, **Data analysis with Kusto Query Language**, and **Analyze monitoring data with Kusto Query Language**

*Used Chat GPT for additional reference on KQL code

*Azure Trainings that were more applicable to our final project are bolded

Step 2: Identifying Dataset and AZURE Trainings Needed

USAID
FROM THE AMERICAN PEOPLE

Identified Our Dataset

- Obtained data from U.S. Agency for International Development
- This data provides supply chain health commodity shipment and pricing data from 2015
- This data is valuable for understanding ranges and trends in pricing, spending, and volumes delivered by country for specific health commodities
- Dataset has 33 Columns and 10335 rows

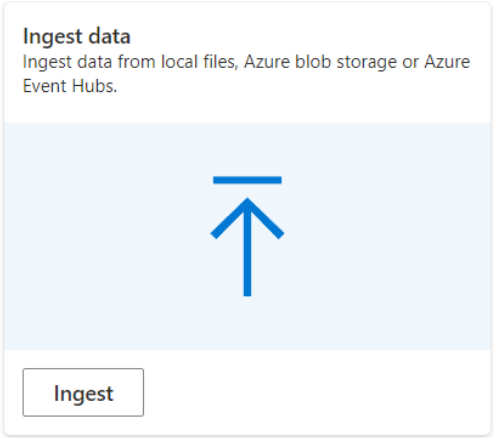
Step 2: Identifying Dataset and Azure Trainings Needed

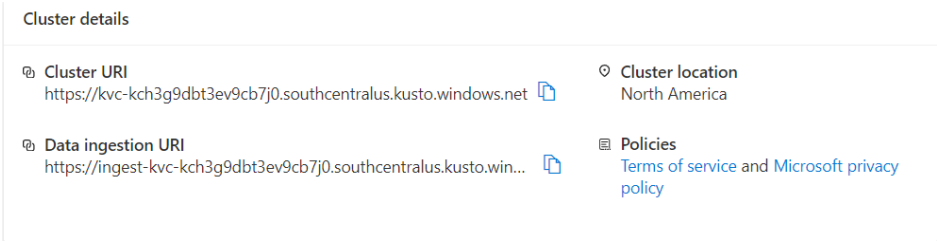
Taking What we Learned From AZURE Trainings

- Learning path: Data Analysis with Kusto Query Language
  - Taught us the fundamentals of query language and use of aggregate functions
    - count, dcount, countif, sum, min, max, avg, percentiles, and others
- Learning path: Data visualization with Azure Data Explorer
  - Taught us how make different graphs and charts

Uploading Our Data

- Uploaded data directly into Data Explorer
  - We discovered that we could upload our data without the use a hot blob or VM
- We ingested our data set as a CSV file from our downloads
- Created a cluster
- Used cluster URI path to create a KQL database

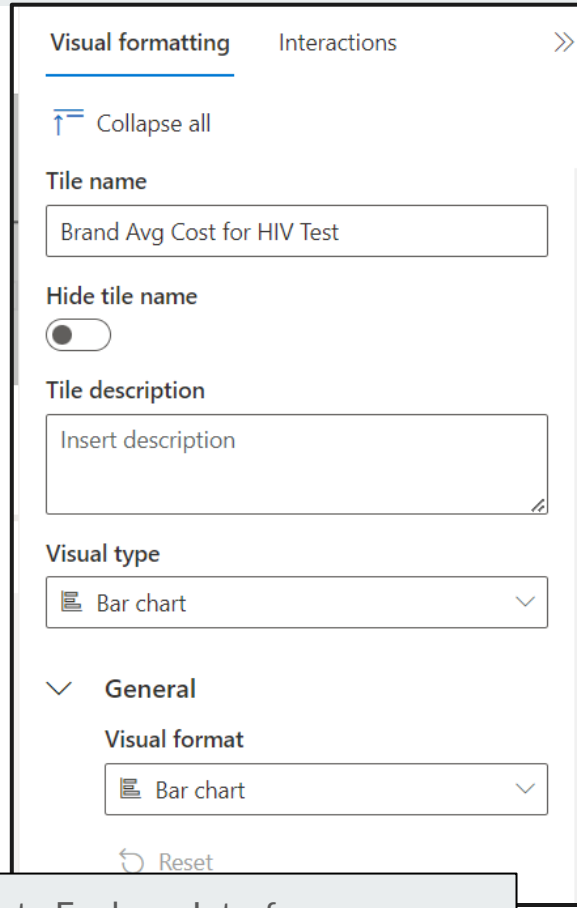




Step 4: Creating Visualizations

Creating Visualisations within Data Explorer

- After uploading the data we were able to begin coding
- Using KQL we were able to narrow our data down into several queries to create different visualizations
    - Works by querying, logs, events, traces, and time series data
    - Uses advanced data statistics for efficient query planning and just-in-time compiled query execution
- After completing our first query to generate a result we selected a visualization type and format
- We used the interface to input the titles and adjust the X and Y scales

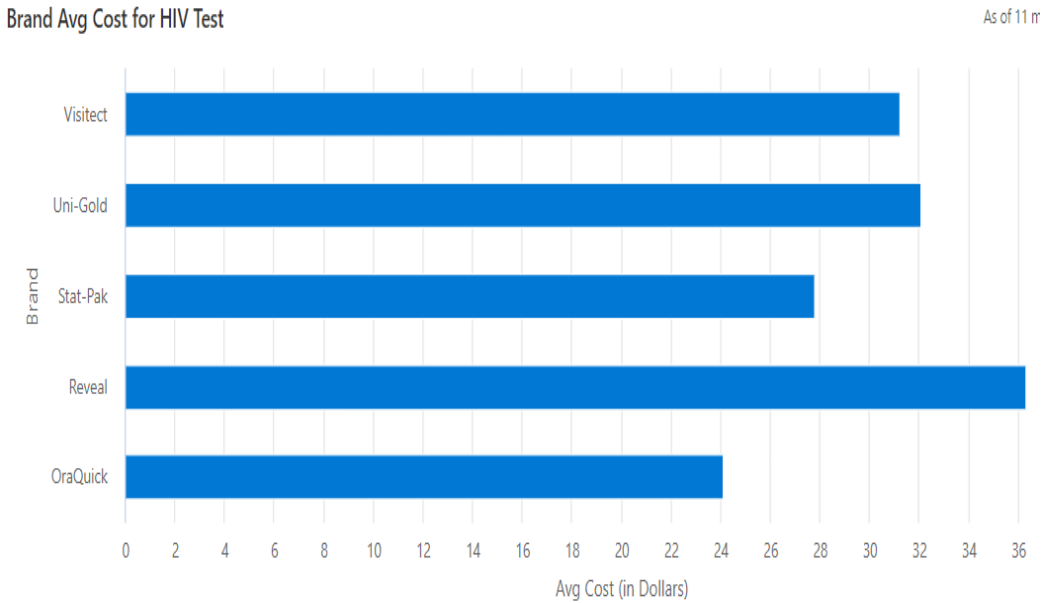


Data Explorer Interface

Our First Visualisation

Displays Top 5 Brands Avg Cost for HIV test kits.

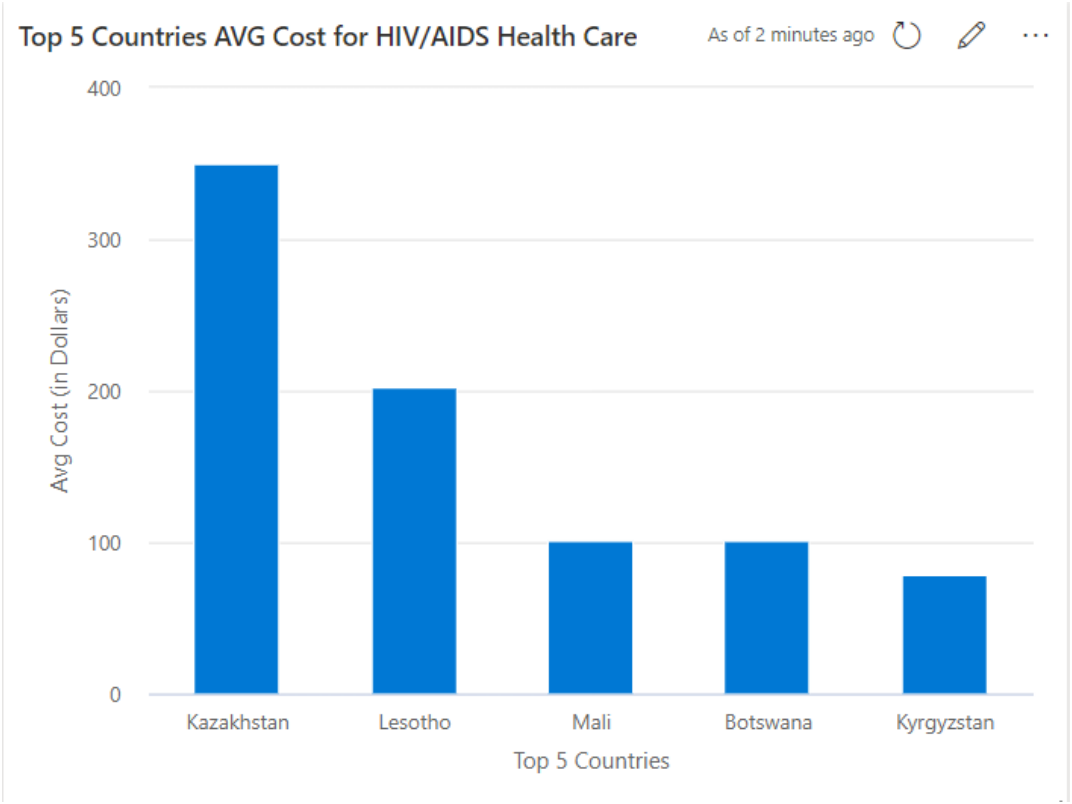
From here we created a dashboard where we would combine all of our future visualizations on one page.



Vis 2: Top 5 Countries AVG Cost for HIV/AIDS Health Care

```
1  ['sales data']
2  |where ['Pack Price'] > 0
3  |summarize
4      AvgCost = avg(['Pack Price'])
5      by Country
6  |sort by AvgCost desc
7  |top 5 by AvgCost
```

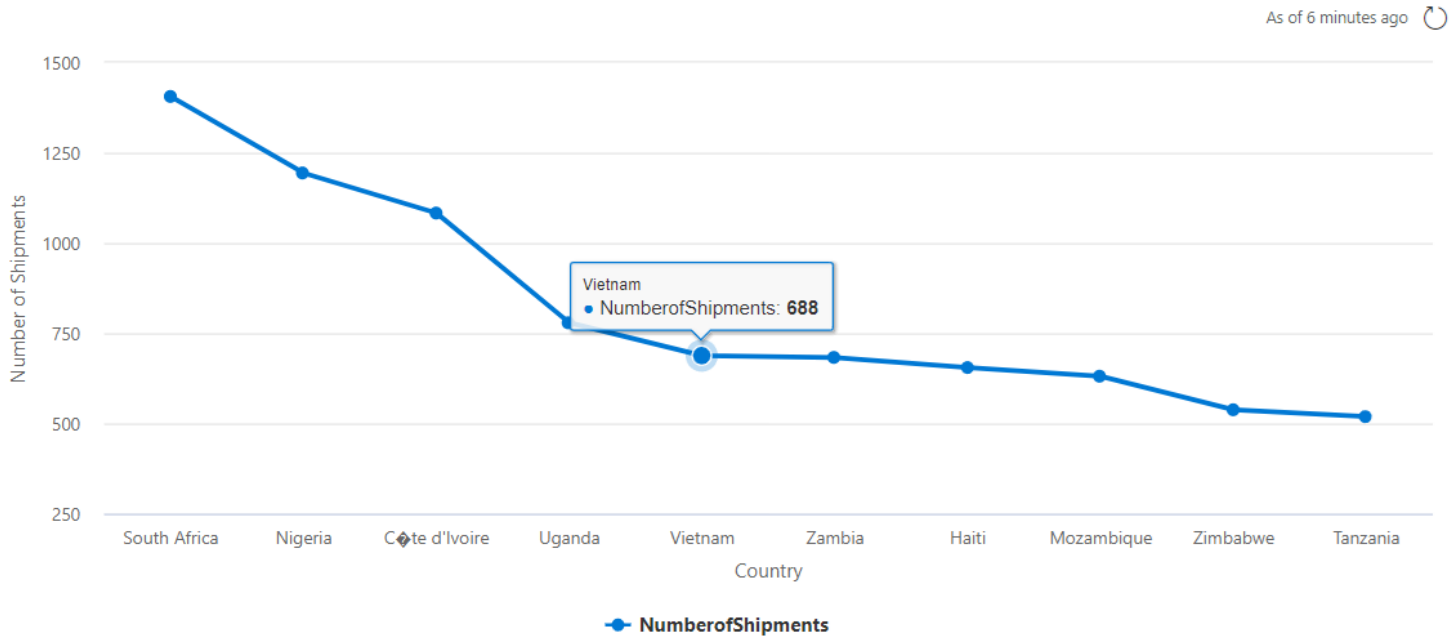
Vis 2: Result


Top 5 Countries AVG Cost for HIV/AIDS Health Care
As of 2 minutes ago
400
300
200
100
0
Avg Cost (in Dollars)
Kazakhstan
Lesotho
Mali
Botswana
Kyrgyzstan
Top 5 Countries

Vis 3: Number of Shipments Per Country

```
1  ['sales data']
2  |summarize NumberofShipments = count(['Shipment Mode']) by Country
3  |top 10 by NumberofShipments
```

Vis 3: Result

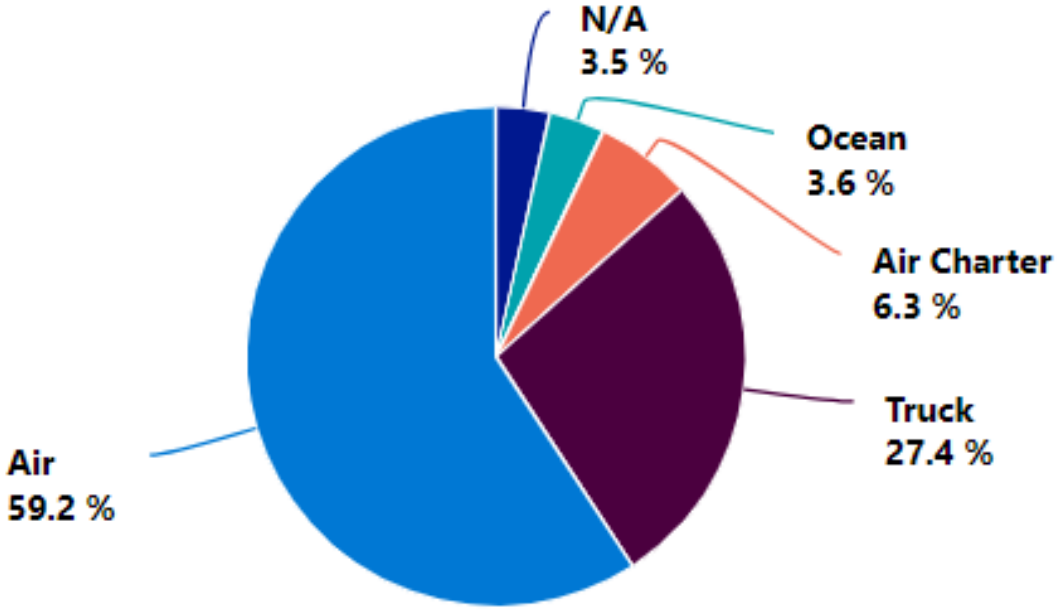

As of 6 minutes ago
1500
1250
1000
750
500
250
Number of Shipments
Vietnam
NumberofShipments: 688
South Africa
Nigeria
C◆te d'Ivoire
Uganda
Vietnam
Zambia
Haiti
Mozambique
Zimbabwe
Tanzania
Country
NumberofShipments

Vis 4: Percentage Modes of Shipment

```
1   ['Sales Data']
2   | summarize
3       COUNTCOST= count(['Pack Price'])
4       by ['Shipment Mode']
5   |sort by COUNTCOST asc
6   |top 5 by COUNTCOST
```

Vis 4: Result

As you can see Air travel made up more than half of the total shipment methods, followed by trucking with slightly over a quarter.



Vis 5: 5 Most Expensive and least Expensive Dosages By Package Prices

Top 5 Code

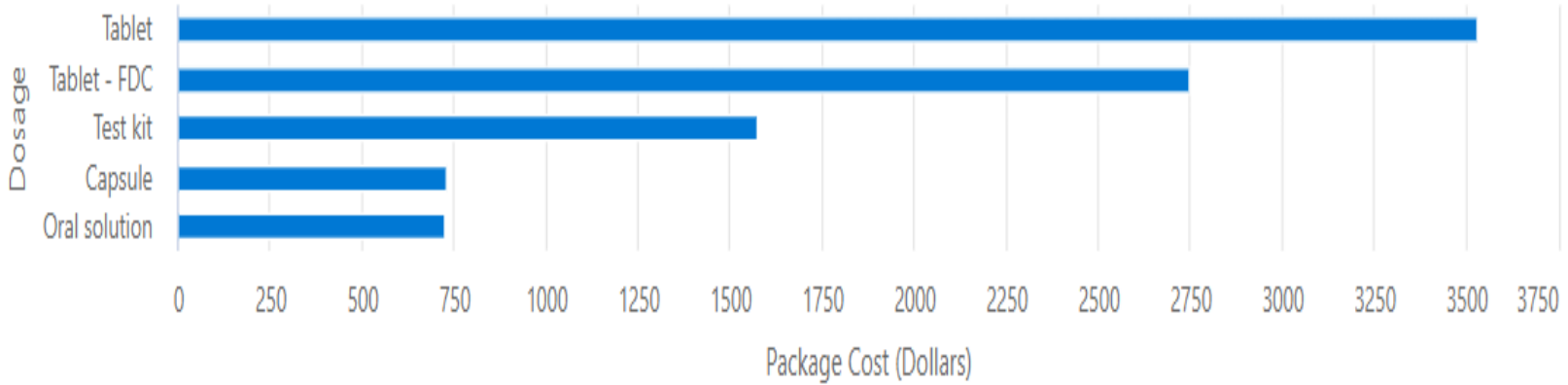
```
['Sales Data']
| summarize
    COUNTCOST= count(['Pack Price'])
    by ['Dosage Form']
|sort by COUNTCOST asc
|top 5 by COUNTCOST
```

Bottom 5 Code

```
['Sales Data']
| summarize
    COUNTCOST= count(['Pack Price'])
    by ['Dosage Form']
|sort by COUNTCOST desc
|top 5 by COUNTCOST asc
```

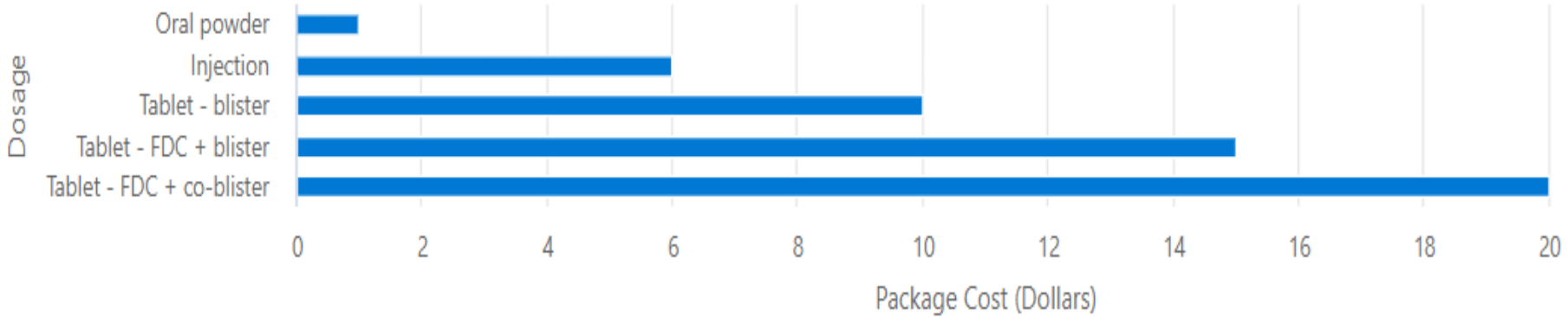
Vis 5: Most Expensive Results

Tablet dosage form is the most expensive package



Vis 5: Least Expensive Results

Oral powder dosage forms ranks as the least expensive



Final Dashboard

Overview

- Azure Trainings guided us towards learning and using KQL and creating visualizations using it
- Discovered Microsoft Data Explorer as a way of uploading data without VM or hot blob
- Were able to begin performing our KQL code from right within Data Explorer
- Visualizations created:
    - Found the most expensive brands of HIV/Aids test kits
    - Determined the countries with the highest cost for HIV/AIDS healthcare
    - Found the 10 countries that had the greatest number of test kits shipped to them
    - Found the percentage each form of transportation was used for shipping test kits
    - Located the 5 least expensive and 5 most expensive forms of dosage treatment
- Final Result of the Dashboard displayed