Carlsberg
Connected Bar

Recommendation System

Enterprise Data Science and Analytics - Bootcamp

M20170998 - Tiago Marques

M20180987 - Joana Ricarte

M20171040 - Sara Matos

Carlsberg
Connected Bar

- IOT and Cloud Architecture
- Big Data Architecture
- Storage and Analytics
- Exploratory Analysis
- Clustering
- Recommendation Systems
- Proof of Concept
- Final Considerations

Develop a recommendation system to:

- Enhance Sales Performance;

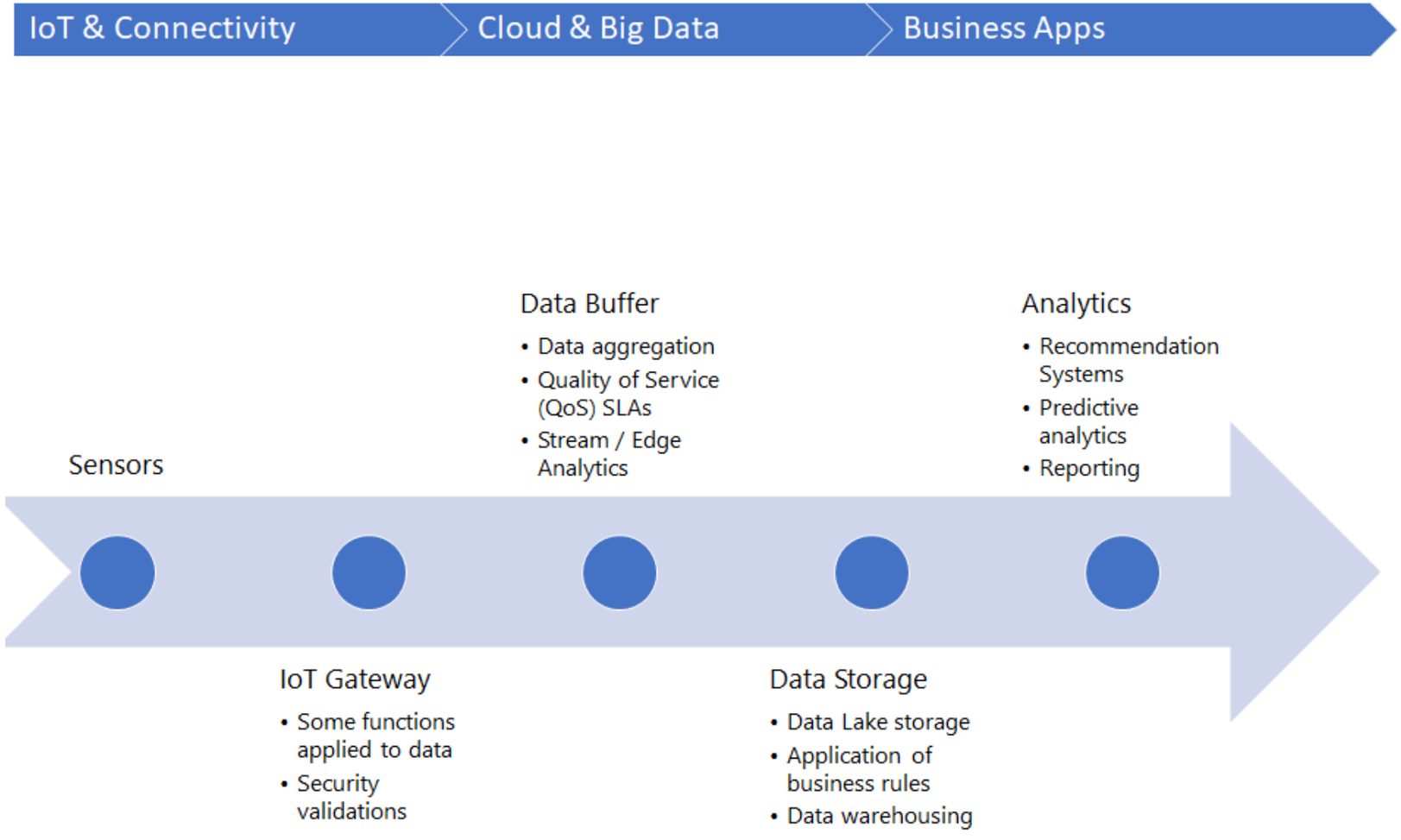
- Provide New and More Realistic Objectives Plan;

- Target and promote new craft beers.

(for Sales Representatives, Customers and Company)

Carlsberg
Connected Bar

- IOT and Cloud Architecture
- Big Data Architecture
- Storage and Analytics
- Exploratory Analysis
- Clustering
- Recommendation Systems
- Proof of Concept
- Final Considerations

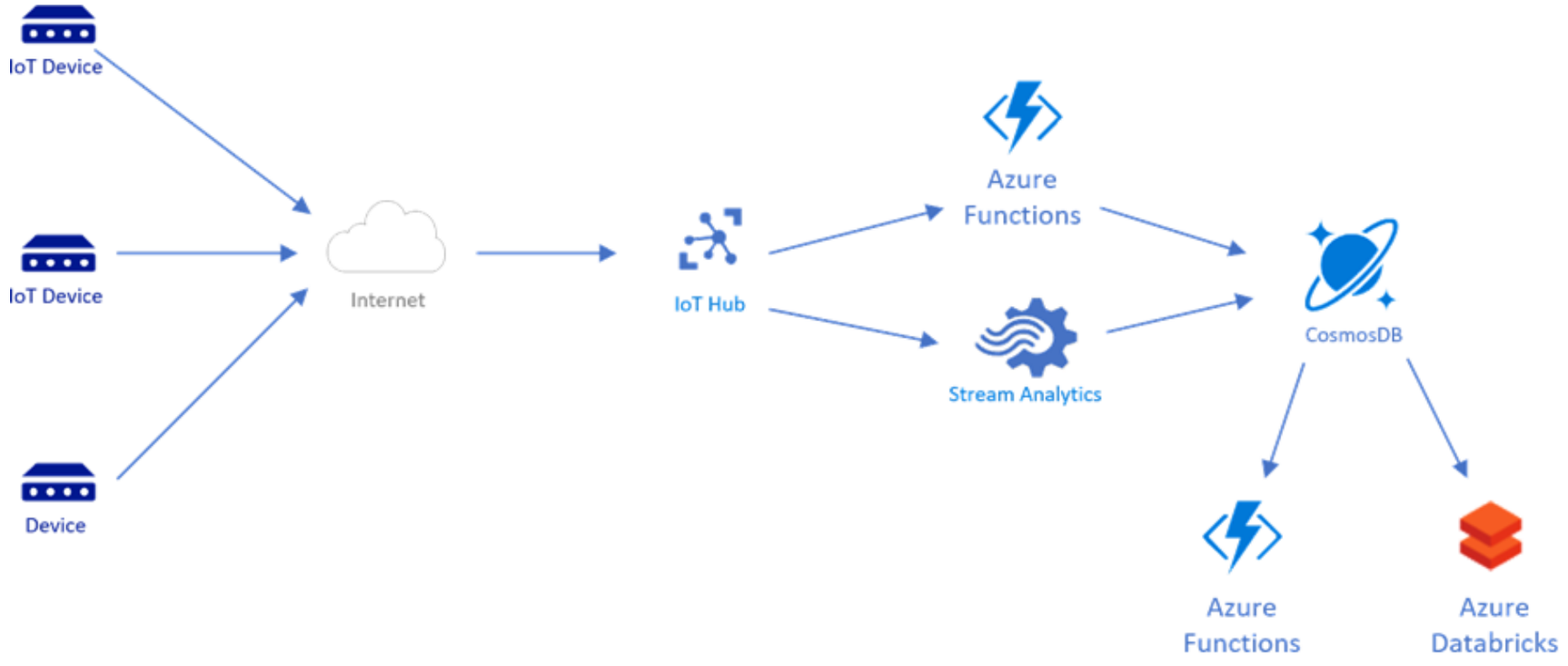


Carlsberg
Connected Bar

- IOT and Cloud Architecture
- **Big Data Architecture**
- Storage and Analytics
- Exploratory Analysis
- Clustering
- Recommendation Systems
- Proof of Concept
- Final Considerations

Our goal:

Simplicity, scalability and application interoperability
over unneeded functionality



Carlsberg
Connected Bar

- IOT and Cloud Architecture
- Big Data Architecture
- Storage and Analytics
- Exploratory Analysis
- Clustering
- Recommendation Systems
- Proof of Concept
- Final Considerations

Big Data Environment for fast Spark development and low maintenance costs

- Reduces the need for cluster management
- Ensures the compatibility of code history and Spark Versions
- Very extensive API for app development
- Integration with a wide variety of data stores and services such as Azure SQL Data Warehouse, Azure Cosmos DB, Azure Data Lake Store, Azure Blob storage, and Azure Event Hub

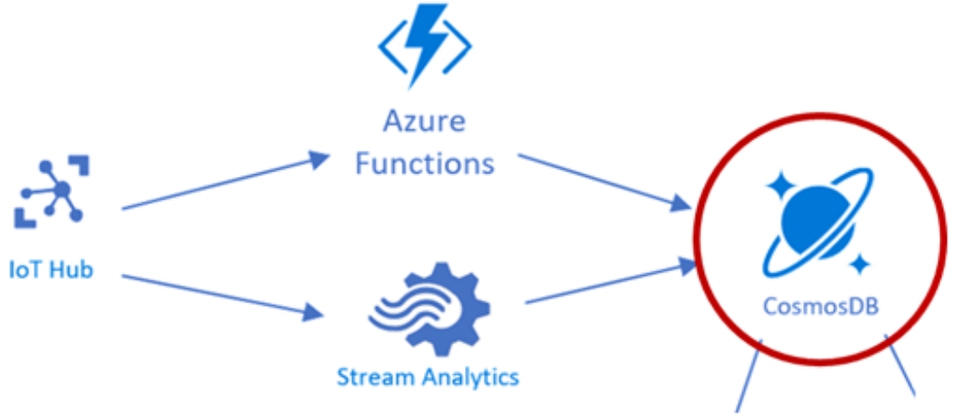
Carlsberg
Connected Bar

In our serverless architecture, Azure Cosmos DB provides the ideal integration to store IoT data. Azure provides a straightforward cloud solutionfor IoT scenarios:

- IoT Hub (getting the data from IoT devices)
- Azure Functions (to apply some kind of logic to data or to identify the database)
- Azure Cosmos DB, to store the data

Cosmos DB in short:

- A globally replicated and scalable database
- A complete SDK available in multiple programming languages
- An API oriented towards modern web app development
- Connectivity to Databricks (both read and write)



Carlsberg
Connected Bar

- IOT and Cloud Architecture
- Big Data Architecture
- Storage and Analytics
- Exploratory Analysis
- Clustering
- Recommendation Systems
- Proof of Concept
- Final Considerations

Power BI is a business analytics service providing interactive visualizations with self-service business intelligence capabilities.
In our context:

- Power BI can turn insights stored in Databricks into easy-to-set visualizations which can be shared across the entire organization
- Clustering and recommendation systems can be used from Databricks
- Power BI can be embedded in web applications, easing the development time and cost of analytical solutions
- Power BI itself is very easily extendable through its marketplace and API
- If needed, Power BI also provides a high-level ELTL framework via its “Query Editor” and the M language

Carlsberg
Connected Bar

- **The majority of the columns are of type String, and those that aren't (last 9 columns) look like they are null for all records except one 436 different sales representatives.**

- **306 sales representatives without (sales_rep) area manager and (sales_rep)region - aprox. 70% of the sales representatives**

- **Only one observation with main_brand and main_brand_id defined**

- **Each Province matches one and only one customer_province. Customer_province is a code for Province?**

- **There are 22 costumer_province without Province**

- **184 Customers don't have an identified Province. of which 122 don't have a customer_providence code 'crl' is an unique identifier of each combination (lat,lon). We can also conclude that 63053 of 68291 costumers don't have a location by latitude and longitude.**

- **Sales information from january 2016 to june 2018**

- **74 un for specialty and craft beers**

- **Negative volume corresponds to returned materials (this does not correspond to our intuition where data is sent by beer tap sensors to our database)**

- **Zero volume corresponds to no sales.**

Carlsberg
Connected Bar

Our approach towards Clustering

We used the K-Means model, which is computationally quite simple and generates satisfactory results

We obtained models for several 2D approaches:

- **Customer sales**
- **Special beers sales by volume | Number of special beers sold**
- **Number of special beer sold in proportion of the total | Volume of special beers sold in proportion of the total**
- **Sales representatives**
- **Special beers sales by volume | Number of special beers sold**
- **Number of special beer sold in proportion of the total | Volume of special beers sold in proportion of the total**
- **Customer geolocation**
- **Number of customers in 1km /10 km radius | Volume of special beers sold**

To obtain the best number of clusters, we tested the information cost for several k-means on smaller samples
To further explore insights about obtained clusters, we put the data in hive tables for each concern of the clustering analysis

Carlsberg
Connected Bar

- IOT and Cloud Architecture
- Big Data Architecture
- Storage and Analytics
- Exploratory Analysis
- Clustering
- Recommendation Systems
- Proof of Concept
- Final Considerations

Our beer recommentation model:

We use a collaborative filtering (customer_id, material_id converted to integers) system based on the Alterating Least Squares method.

Since we don't have a "rating" in the data provided to us, we use the proportion of C&S beers sold over the total volume, we then multiply by 10 and round to no decimal cases, thus obtainind a 0-10 rating grid.

We used two different measures for prediction accuracy: RMSE and MSE.

The predictions obtained are then joined with original data and exported to a hive table for further exploration in Power BI

Carlsberg
Connected Bar

- IOT and Cloud Architecture
- Big Data Architecture
- Storage and Analytics
- Exploratory Analysis
- Clustering
- Recommendation Systems
- **Proof of Concept**
- Final Considerations

Consumption Dashboard

Overview

Sales Rep.

Customers

Beer | Volume (hl)

13,76K !

Goal: 14097 (-2.39%)

Wine | Volume (hl)

610,00 !

Goal: 723 (-15.63%)

Special Craft Beer | Volume (hl)

2,69K ✓

Goal: 2564 (+5.03%)

TOP clients

customer	Average of volumehl
OSCELLA S.p.A.	10,74
ALPORI FESTA & C. SPA	10,68
S.MARCO BEVERAGE NETWORK	9,22
NUOVA ALBA 82 S.r.l.	9,20
	0,83

TOP Sales Representative

sales_rep	Average of volumehl
SAUSA SERGIO	21,64
CERINO LUIGI	11,83
DIREZIONALE CATERING	8,00
ASSIRELLI FAUSTO	7,41
BIANCHINI MARCO	6,00

esri THE SCIENCE OF WHERE™

Special Crafts Dashboard

Overview

Sales Rep.

Customers

Nr of Special Crafts Being Sold

31

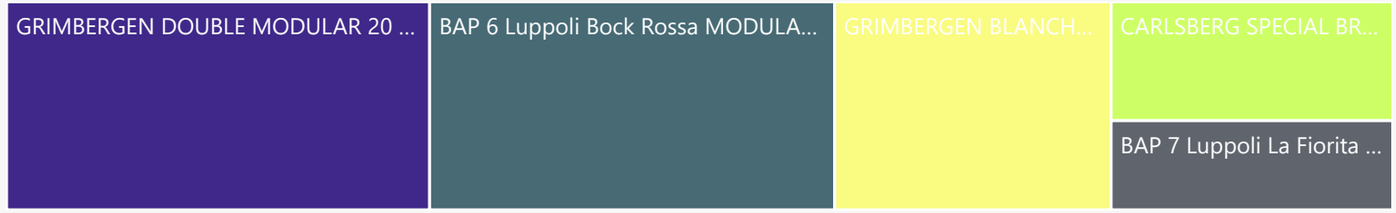
Goal: 27 (+14.81%)

Volume of Special Crafts Being Sold

2,06

Goal: 1,62 (+26.88%)

More popular Crafts



Beers sold and Recomended actions

material	# customers	Recommendation
BAP 7 L'Affumicata MODULAR 20	337	Promote
BROOKLYN NARANJITO MODULA...	121	Promote
TUBORG RED MODULAR 20 fusto...	128	Promote

material	# customers	Recommendation
ANGELO BROWN ALE MODULAR ...	88	Maintain
BAP 7 L'Esotica MODULAR 20	41	Maintain

material	# customers	Recommendation
ANGELO BLANCHE modular20 fs ...	3	Discontinue
BAP 4 Lager BIO MODULAR 20		Discontinue
BAP 7 Luppoli L'Estiva MODULAR ...	48	Discontinue

TOP Sales Representative

sales_rep	Average of volumehl
CAPONE SANDRO	2,44
ASSIRELLI FAUSTO	2,39
TOSTI MARCO	2,34
BIANCHINI MARCO	2,27
CINIERI DONATO	1,95
FENDONI MIRKO	1,64
SORIANO LUIGI	1,47
GOBBI BARBARA	1,39
BONINO CLAUDIO	1,29
ZENOBI CARLO	1,20

lat, lon and volumehl

ABAGNALE ROBERTO

GENTILI STEFANO

Overview

Sales Rep.

Customers

Beer | Volume (hl)

6,00 !

Goal: 8 (-25%)

Wine | Volume (hl)

1,00 ✓

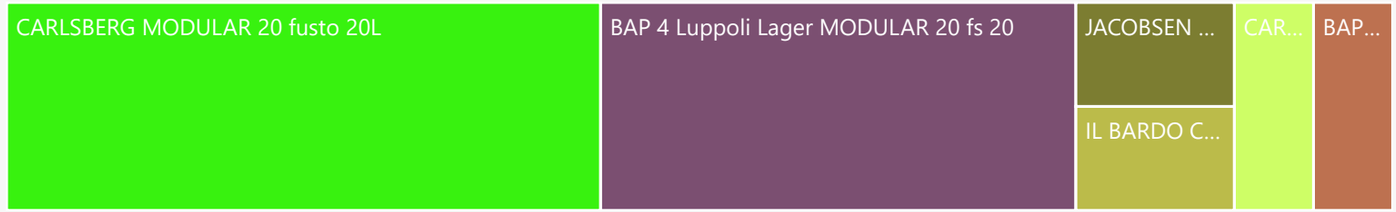
Goal: 0

Special Craft Beer | Volume (...

0,00 !

Goal: 3 (-100%)

More popular Beers



Beers sold and Recomended actions

material	# customers	Volume (hl)	Reference	Recommentation
BROOKLYN LAGER MODULAR 20 ...	1	1,00	0,53	Maintain

material	# customers	Volume (hl)	Reference	Recommentation
BAP 4 Luppoli Lager MODULAR 2...	3	12,00	5,26	Promote
TUBORG FLEX 20 fusto 20L	1	1,00	4,41	Promote
BAP 4 Luppoli Lager FLEX 20 fS 20L	1	2,00	4,24	Promote

material	# customers	Volume ...	Reference	Recommentation
IL BARDO GARGANEGA GLERA M...			15,00	Discontinue
TUBORG MODULAR 20 fusto 20L			3,89	Discontinue
TUCHER UBERSEE MODULAR 20 f...			3,73	Discontinue

Must Try

.

BAP 6 Luppoli Bock Rossa MODULAR 20 fs

BAP 7 Luppoli La Fiorita MODULAR 20 fs20

GRIMBERGEN DOUBLE MODULAR 20 fusto 20L

esri THE SCIENCE OF WHERE™

OSCELLA S.p.A.

Overview

Sales Rep.

Customers

Beer | Volume (hl)

486,00 ✓

Goal: 159 (+205.66%)

Wine | Volume (hl)

60,00 !

Goal: 66 (-9.09%)

Special Craft Beer | Volume (...

102,00 ✓

Goal: 25 (+308%)

DOMODOSSOLA

More popular Beers

Beers sold and Recomended actions

material	# customers	Volume (hl)	Reference	Recommentation
BAP 4 Luppoli Lager MODULAR 2...	1	594,00	5,26	Maintain
BAP 7 L'Affumicata MODULAR 20	1	11,00	0,15	Maintain

material	# customers	Volume (hl)	Reference	Recommentation
GRIMBERGEN BLONDE MODULA...	1	3,00	0,85	Promote
BROOKLYN LAGER MODULAR 20 ...	1	1,00	0,53	Promote

material	# customers	Volume ...	Reference	Recommentation
IL BARDO GARGANEGA GLERA M...			15,00	Discontinue
KRONENBOURG 1664 FLEX 20 fus...			2,44	Discontinue
TUCHER PILSNER MODULAR 20 f...			2,15	Discontinue

Must Try

.

BAP 6 Luppoli Bock Rossa MODULAR 20 fs

GRIMBERGEN BLANCHE MODULAR 20 fusto 20L

GRIMBERGEN DOUBLE MODULAR 20 fusto 20L

esri THE SCIENCE OF WHERE

Carlsberg
Connected Bar

- IOT and Cloud Architecture
- Big Data Architecture
- Storage and Analytics
- Exploratory Analysis
- Clustering
- Recommendation Systems
- Proof of Concept
- Final Considerations

Conclusion

The possibility of working with fresh IoT data is one of the most interesting fields in analytics.
This includes the possibility of having near-realtime insights on data.

Nowadays, cloud providers have fully integrated and secure ecosystems to work, but the number of solutions and the cost must be well calculated in function of the business needs.

We opted for a fairly simple architecture, that provides a solid basis for developers to build a front-end application and allows a quick time-to-market.

On the data science side, Azure Databricks and Power BI turned to be a powerful combination to quickly extract insights from data and even make recommendations on it.

The data that we got had some gaps, but it was interesting to foresee a lot of applications.
As we expected to get fresher data, it will become possible to apply other algorithms, such as:

- **time series algorithms on shorter periods of time**
- **better collaborative filtering**
- **more meaningful clustering analysis**