

Introduction to NoSQL Databases

A business case for efficient Gourmet Food delivery



Leila Bagheri

Noah Prozan

Shalini Chawla

Data Science Team @ Acme Gourmet Meals

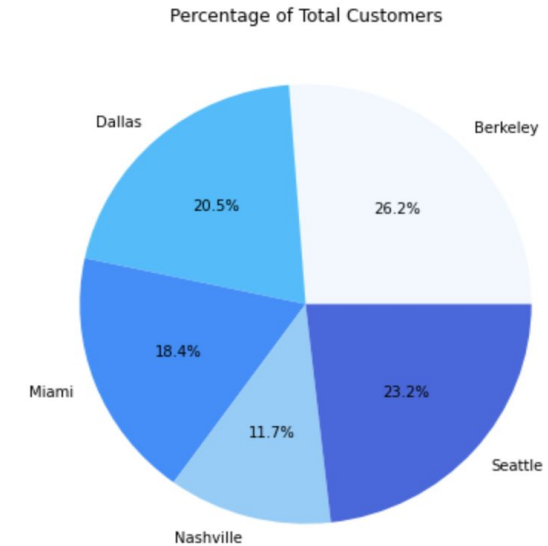
Where we are today? Business

Business Standpoint

- 5 stores in US: healthy business
- Berkeley store currently serving the San Francisco Bay Area is our best store
- Customers come to the store

Business Standpoints

- Increase our customer base in strategically important locations
- Add more pick up locations
- Try out new business models
 - Reach the customers where they are

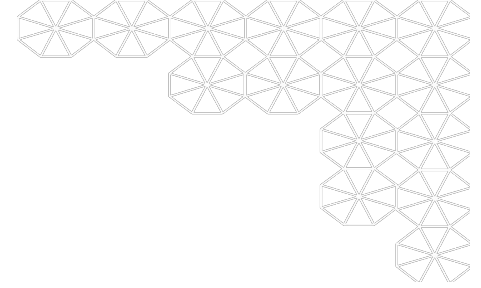


BART

- 50 stations throughout the Bay Area
- People are coming back to work from the Pandemic
- An efficient network of delivery/pickup points
- Ability to connect at different points in a region to see where our product is in biggest demand



Databases: where are we today?



Database Standpoint

- We want best in class data solutions in order to maximize business performance, have existing relational database that covers customers, stores, line items, etc.
- Performance issues as we continue to scale our business. How to we add to our existing data stack?

Database Solutions

- NoSQL databases: not only SQL
 - Scalability
 - Speed: fast, data stored per need
 - Flexibility: planning needs to be done in relational databases
 - Natural fit: different problems to handle for our business needs

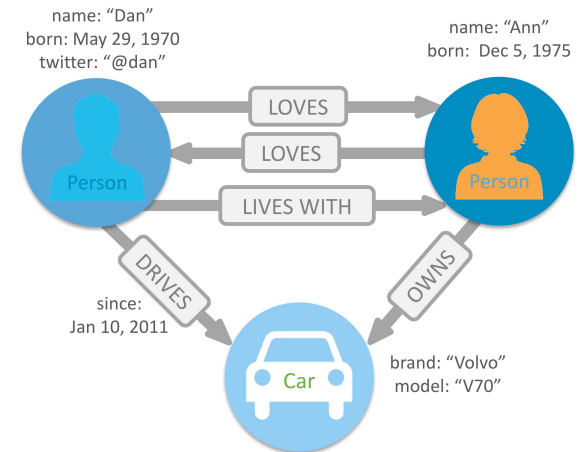
We will give a detailed overview of the BART case enhanced by the NoSQL databases and how they can improve delivery network design, and real time inventory tracking/ analytics.

Neo4j - Graph Database



Business Case: Identify optimal delivery points

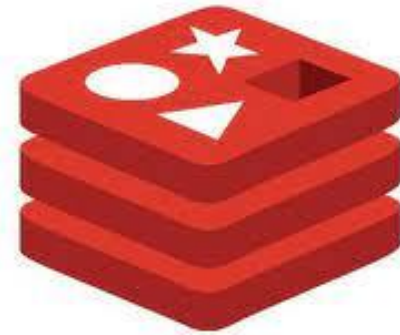
- Nodes and relationships instead of tables
- Ability to handle complex connections between the data points
- Relationships are stored as part of data
- Natural fit for the BART infrastructure of stations–lines–travel times
- Prebuilt support for
 - Path Finding – Optimum delivery routes delivery routes
 - Analytical decision making on inventory
 - Community detection – identify groups of nodes
 - Targeted marketing decisions?
 - Centrality – Importance of distinct nodes in the network
 - Key customers



Redis: Key-Value Data Store

Business Case: Live Inventory Tracking

- Fast, in-memory, key-value data store
- Performance
- High availability and scalability
- Ease of use



MongoDB: Document Database

Business Case: Advanced Analytics

- Flexible document schemas
- Powerful querying and analytics
- Easy horizontal scale out
- High performance

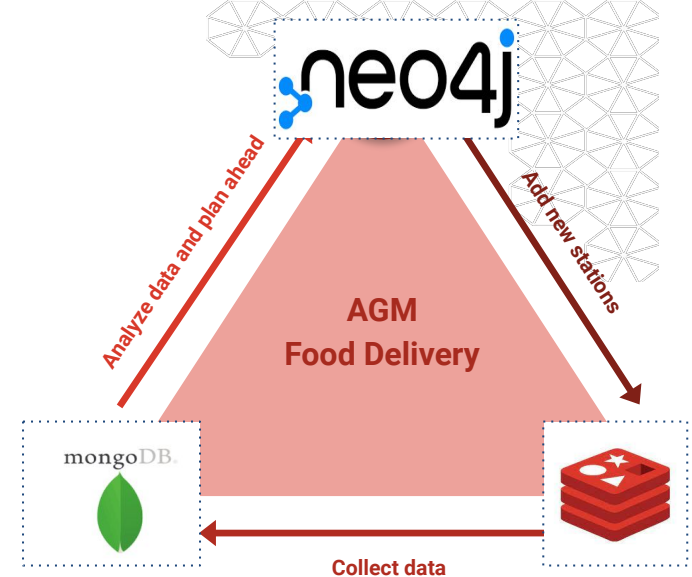
mongoDB.



Final Proposed Solution

Step I: Neo4j to identify stations

- Build a graph database in Neo4j for the BART system
- Use the Louvain Modularity Algorithm to identify strong knit communities of stations
- From each community, pick one station that has the potential to reach the maximum customer base (Using geodesic population grids)
- Use the selected 11 stations as pilot pickup points for Phase I
- At each iteration, we will be coming back to Neo4j to identify the most efficient extension points to our delivery network, which may be driven by additional criteria like shortest paths from our delivery points, closeness to other stations, degree of betweenness of a particular station



Final Proposed Solution continued...



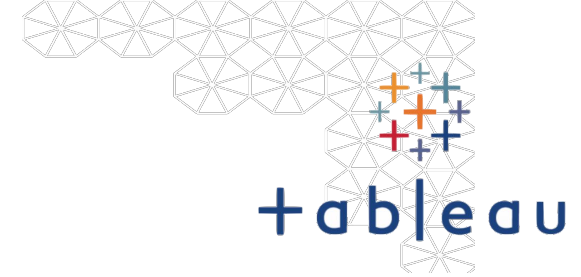
Step II: Redis to track live inventory

- Use the key-value store to track product inventory levels at each delivery point station
- Use this information to plan number of deliveries and the quantities per product to be delivered each day to the delivery points

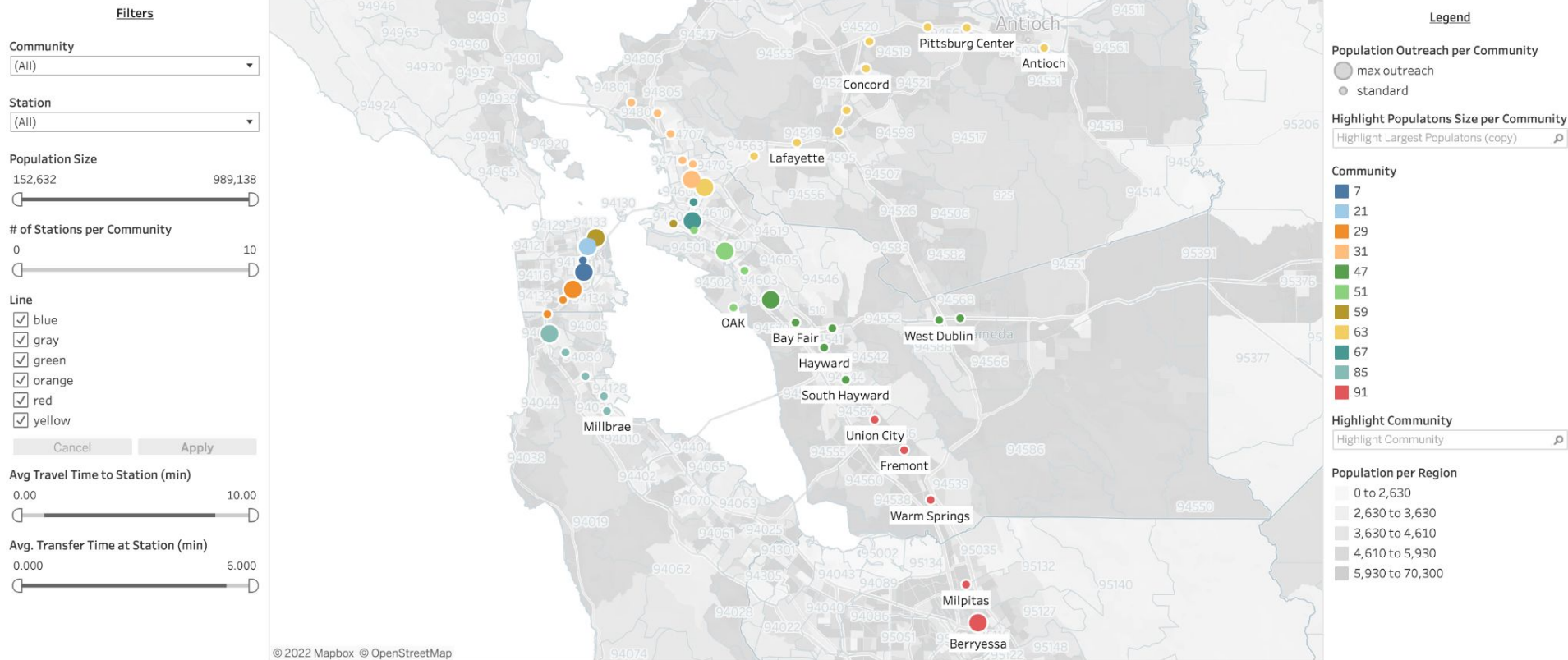
Step III: MongoDB for analytics

- Use the document store to store information in different point of views for advanced analytics
- Information can be stored for a product POV, delivery location POV, weekday POV.
- Use data to evaluate success/failure/limitations of delivery points and make better informed business decisions at each iteration

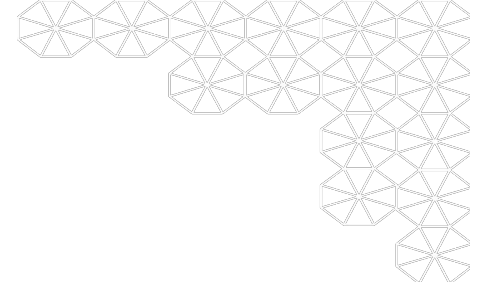
Visualization Demo



Bart Lines by Community



Timeline



Phase 1

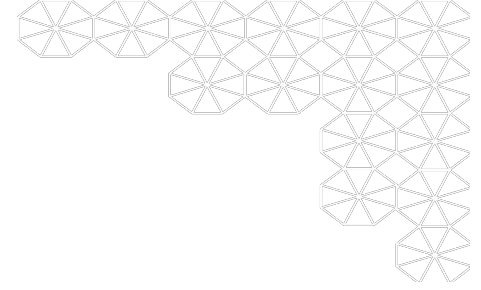
- **Month 1:** Launch food delivery pickup locations at 11 Bart Stations within each community:
 - Rockridge, Ashby, 19th St, Fruitvale, San Leandro, Berryessa, Montgomery St, Civic Center, 24th St Mission, Glen Park, Colma
- **Month 2–7** (6 month window period): Gather performance data from each station on population outreach and popularity, maximize inventory strategy

Phase 2

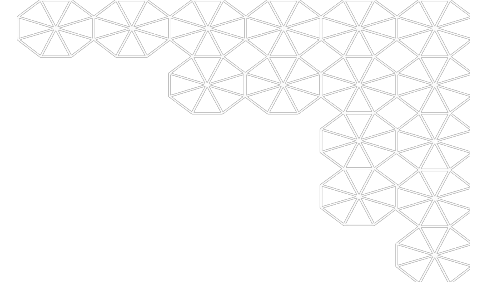
- **Month 8:** Limit to top 3 communities based on performance:
 - Allocate stronger resources & more traction to those top 3 communities
- **Month 9–12** (4 month window period): Gather performance data from Phase 2 to pinpoint a location for where to launch our second store

Phase 3

- **Year 2:** Planning period for store openings nationwide from Store 2 performance



Questions?



References

- <https://neo4j.com/>
- <https://www.mongodb.com/>
- <https://redis.com/>
- https://en.wikipedia.org/wiki/Bay_Area_Rapid_Transit

Deliverable Links

- [Tableau Visualization](#)
- Github