**Team 12**

Dingruo Yang, Muyu Yan, Phoebe Chang,

Siva Kallur, Stephen Champagne, Sushanth Gaddam

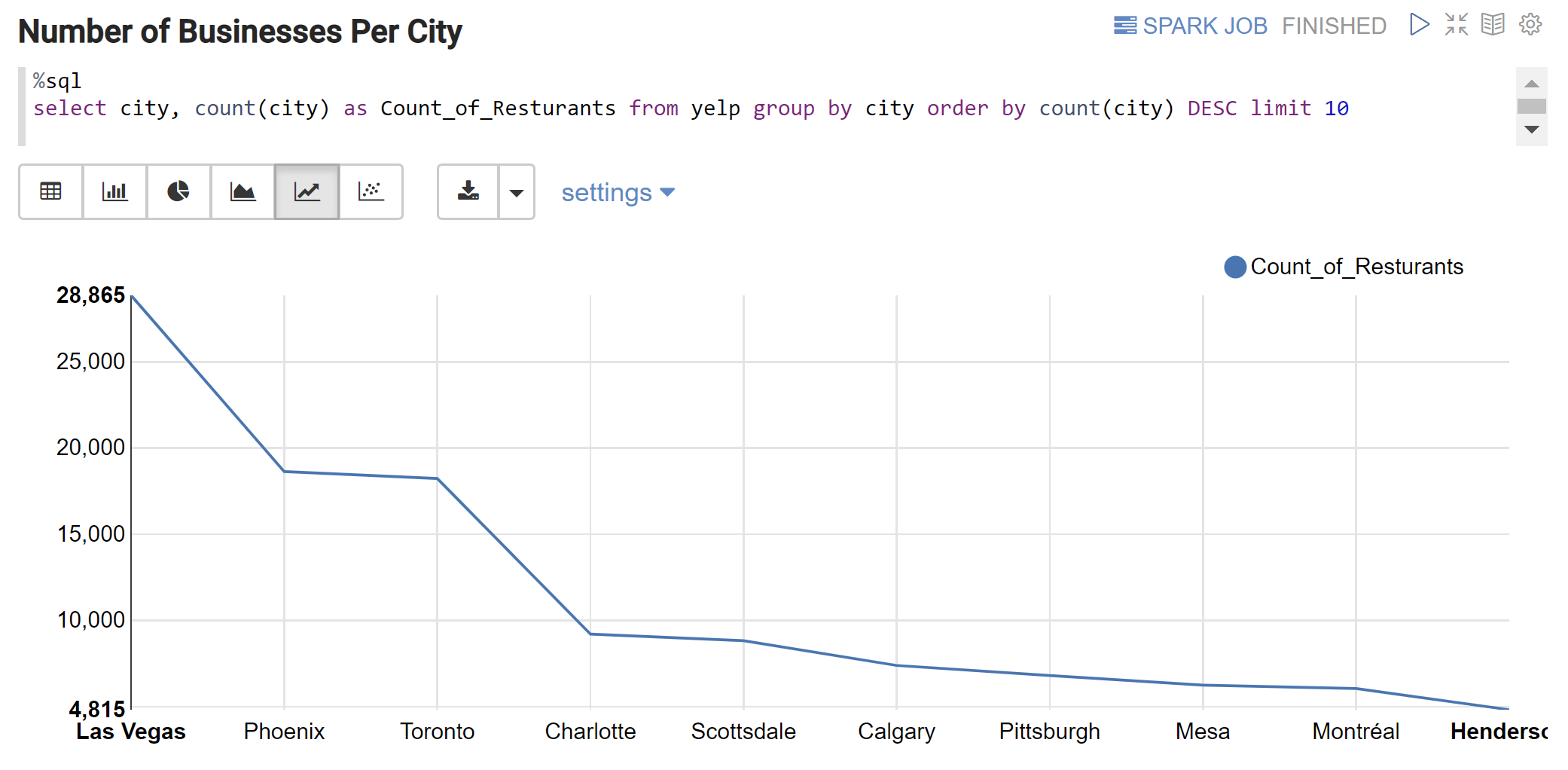
Shareable Exploratory Data Analytics with Python, SQL, or Spark using Apache Zeppelin

December 10, 2018

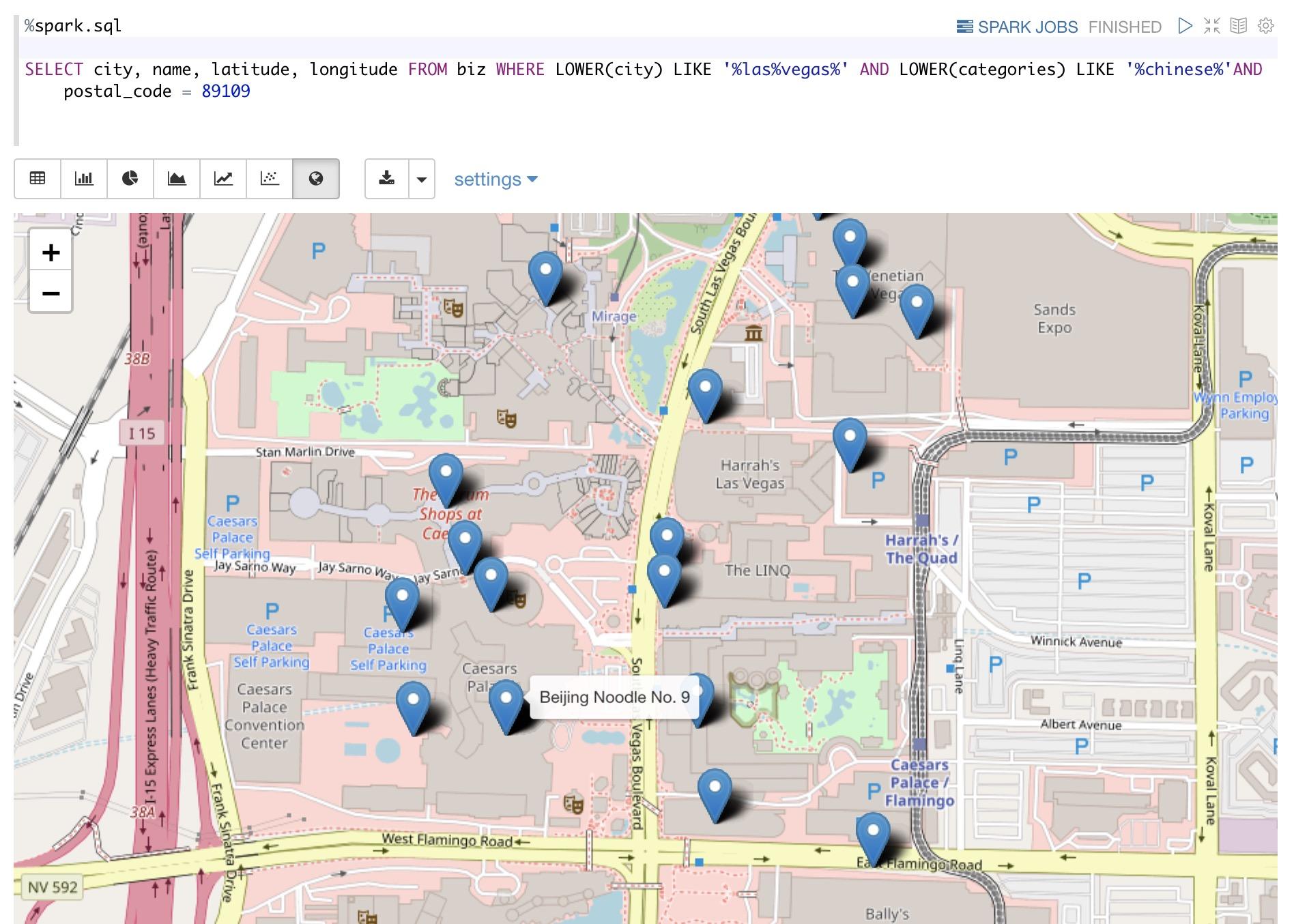
# Overview

Apache Zeppelin is a web-based notebook system for sharing interactive data analytics. It currently supports various languages such as Python, SQL, and Spark (many others as well). These notebooks can be shared very easily using a link and additional analytics can be performed on the shared notebook. Below are reasons you should switch your Exploratory Data Analytics to Apache Zeppelin.

|  |  |  |
| --- | --- | --- |
| **Easy to Use**  Expertise in Python? SQL? Spark? Any of these languages can be used while using Zeppelin notebooks! | **Easy to Share**  Each analysis is linkable and can be individually sent. Check our analysis on exploring Yelp review data in the QR Codes. | **Open Sourced**  Don’t feel like paying for Tableau anymore? Get started by downloading **free** Zeppelin here: <https://zeppelin.apache.org/> |

Exploring Yelp Data to Find Locations and Categories for a New Restaurant

**1.** SQL queries can be run and converted automatically to a variety of graphs that come with Zeppelin. We analyze Las Vegas Restaurants since it has the highest quantity.



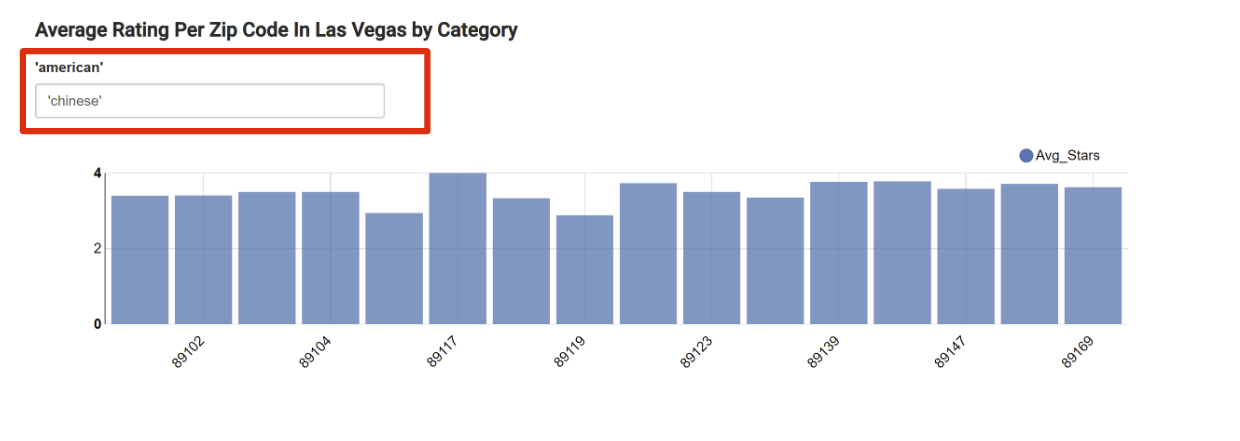
**2.** Chinese Restaurants appear to have the lowest rating across all categories. This could be a good opportunity! Check out how many restaurants nearby and where could be the best location.



**3.** Take a further look at how customers care when giving positive or negative feedback. Seems like must be careful about the taste of dim sums!



**4.** Using correlation matrix to discover the most relevant features of restaurants via AWS Jupyter Hub. We found that restaurants that offer delivery, outdoor seating, take-away, reservation services are most correlated to rating and review counts, these are probably good features for our restaurant.

**5.** Zeppelin also allows you to create interactive dashboards. Load the dashboard up and try for yourself!

Siva Kallur - [kallu016@umn.edu](mailto:kallu016@umn.edu), Sushanth Gaddam - [gadda010@umn.edu](mailto:gadda010@umn.edu), Dingruo Yang - [yang6548@umn.edu](mailto:yang6548@umn.edu) Stephen Champagne - [champ183@umn.edu](mailto:champ183@umn.edu), Muyu Yan - [yan00108@umn.edu](mailto:yan00108@umn.edu), Phoebe Change - [chan1760@umn.edu](mailto:chan1760@umn.edu)