Yelp Recommender System

... Recommendations Tailored to Your Taste

Not All 5-stars Are Created Equal

Recommender system is used in many e-commerce as well as service subscription platforms to help users discover new products by recommending either similar products/ things similar users have used/liked.
To users, these systems help you discover To businesses, it helps drive engagement as well as cross/up-sale.
Cross/Up-Sell (e.g. Amazon) | Explore (e.g. Spotify, Netflix)

Will I like this place with such glowing reviews?

Can I find new, reliable recommendations?

How can Yelp attract more users and businesses?

What is a recommender system

Collaborative Filtering

- Recommend based on past behaviours
- Assume similar users like similar things
- Item-item based
- User-user based
- **Matrix factorization**

Content-based Filtering

- Recommend based on past behaviours
- Assume users like similar items

Hybrid

Combine previous approaches

Matrix Factorization

	Popeyes	McDonald's	The Senator
Alice	?	4	3
Bob	2	?	4
Tim	5	4	3

	Fast Food	Chicken	Ambience
	Hidden Feature 1	Hidden Feature 2	Hidden Feature 3
Alice	0.3	0.4	0.3
Bob	0.1	0.2	0.5
Tim	0.8	0.6	0.4

- Hidden features are hidden! Meant to help imagine backend process
- A matrix can be factorized into two matrices.

A (i rows x j cols) = M (i rows x k cols) . N (k rows x j cols)

		Popeyes	McDonald's	The Senator
Fast Food	Hidden Feature 1	0.5	1.5	0
Chicken	Hidden Feature 2	1.8	0.4	0.4
Ambience	Hidden Feature 4	0.1	0.1	1.3

Matrix Factorization & Machine Learning

	Popeyes	McDonald's	The Senator
Alice	?	4	3
Bob	2	?	4
Tim	5	4	3

	Fast Food	Chicken	Ambience
	Hidden Feature 1	Hidden Feature 2	Hidden Feature 3
Alice	0.3	0.4	0.3
Bob	0.1	0.2	0.5
Tim	0.8	0.6	0.4

Iterative process to update users and restaurants to minimize Error

Predict ? using final matrices

Alice – McDonald's Rating

0.64 = 0.3*1.5 + 0.4 * 0.4 + 0.3 * 0.1

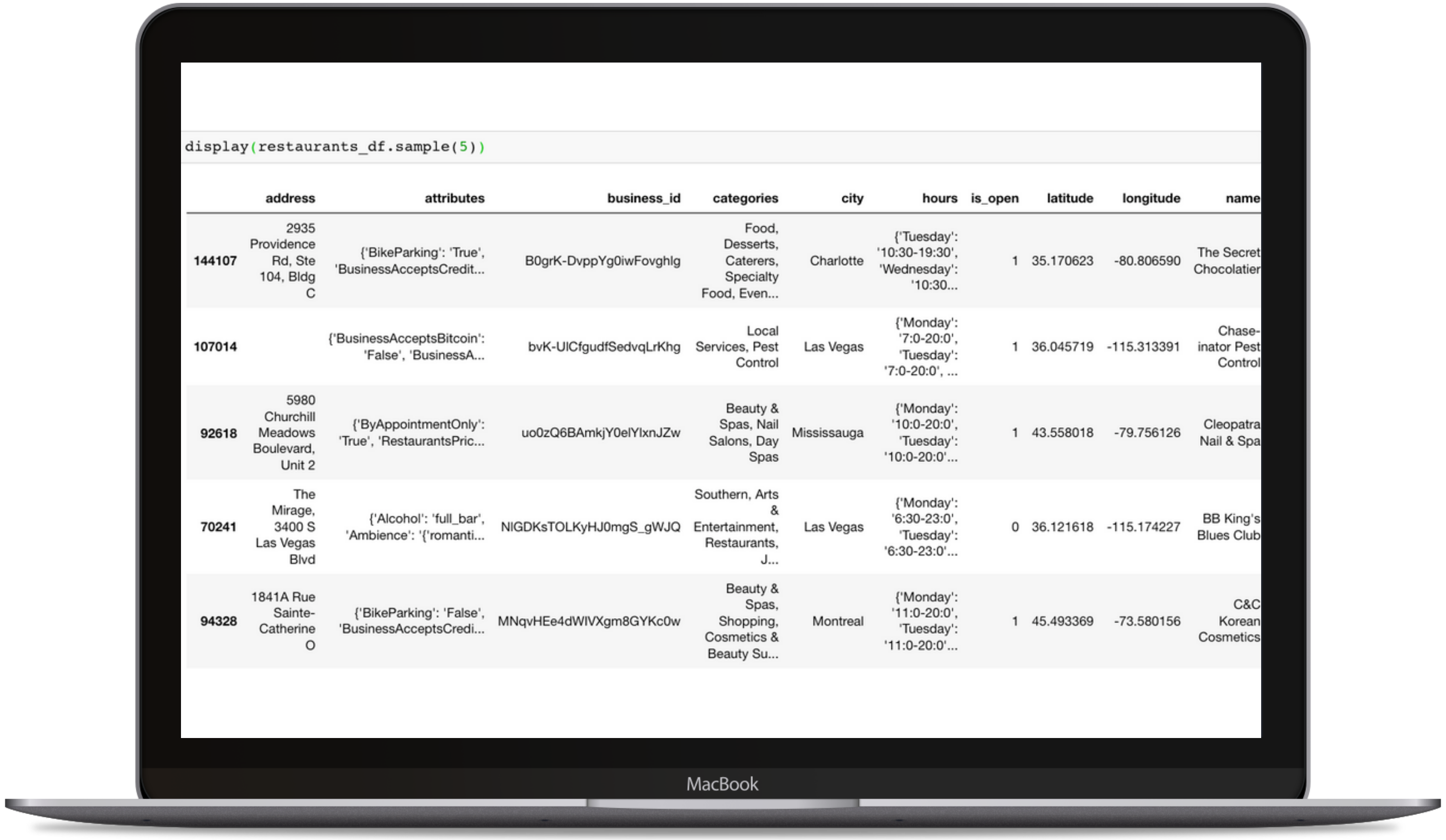
Error

3.36 = 4 - 0.64

		Popeyes	McDonald's	The Senator
Fast Food	Hidden Feature 1	0.5	1.5	0
Chicken	Hidden Feature 2	1.8	0.4	0.4
Ambience	Hidden Feature 4	0.1	0.1	1.3

Where It All Started

Data Source, Format, Structure



```
display(restaurants_df.sample(5))
```

	address	attributes	business_id	categories	city	hours	is_open	latitude	longitude	name
144107	2935 Providence Rd, Ste 104, Bldg C	{'BikeParking': 'True', 'BusinessAcceptsCredit...	B0grK-DvppYg0iwFovghlg	Food, Desserts, Caterers, Specialty Food, Even...	Charlotte	{'Tuesday': '10:30-19:30', 'Wednesday': '10:30...	1	35.170623	-80.806590	The Secret Chocolatier
107014		{'BusinessAcceptsBitcoin': 'False', 'BusinessA...	bvK-UlCfgudfSedvqLrKhg	Local Services, Pest Control	Las Vegas	{'Monday': '7:0-20:0', 'Tuesday': '7:0-20:0', ...	1	36.045719	-115.313391	Chase-inator Pest Control
92618	5980 Churchill Meadows Boulevard, Unit 2	{'ByAppointmentOnly': 'True', 'RestaurantsPric...	uo0zQ6BAmkjY0elYlxnJZw	Beauty & Spas, Nail Salons, Day Spas	Mississauga	{'Monday': '10:0-20:0', 'Tuesday': '10:0-20:0'...	1	43.558018	-79.756126	Cleopatra Nail & Spa
70241	The Mirage, 3400 S Las Vegas Blvd	{'Alcohol': 'full_bar', 'Ambience': '{'romanti...	NIGDKsTOLKyHJ0mgS_gWJQ	Southern, Arts & Entertainment, Restaurants, J...	Las Vegas	{'Monday': '6:30-23:0', 'Tuesday': '6:30-23:0'...	0	36.121618	-115.174227	BB King's Blues Club
94328	1841A Rue Sainte-Catherine O	{'BikeParking': 'False', 'BusinessAcceptsCredi...	MNqvHEe4dWlVXgm8GYKc0w	Beauty & Spas, Shopping, Cosmetics & Beauty Su...	Montreal	{'Monday': '11:0-20:0', 'Tuesday': '11:0-20:0'...	1	45.493369	-73.580156	C&C Korean Cosmetics

- Yelp-published data on Kaggle.com
- 1.5 million users | 6 million reviews | 155k businesses
- Files in JSON (think dictionary with word-definition pairs)

What I Need to Do

Analyze

What constitute similar taste?

- Previous Ratings of the Same Restaurants
- Preferences ($ range, Cuisines, locations of visits)
- Language In Text Reviews

Prepare

Extract **users' preferences for price range, cuisines**

Create a big table of **users' reviews for all Yelp businesses**

Create groups of **similar users based on custom profiles**

Model

Test clustering -> matrix factorization vs matrix factorization only

Generate possible ratings of unrated businesses for each users

Lessons Learnt

Experiments do not always work

Create a cluster and feed specific cluster into matrix factorization. The model performed worse!

Wrong direction with natural language processing

Analysis Paralysis

Strike a balance between creating a first iteration vs exploring every corner of a complex problem

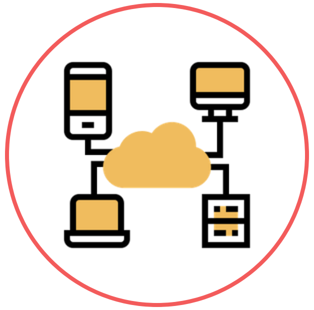
Real-world Mindset

Discover my new area of interest - Recommender System!

A nice blend of data science and engineering to solve real world challenges

Future Direction

Explore Hybrid Approach



Database to Speed Up

Neo4J - Social Graph

Question?