



Where Should New Grads Move?

Maxine Gardner, Kedar
Godbole, Sonali Hornick,
Kimble Horsak, Deeksha
Koonadi, Luke Leon

MSBA Fall 2024: Analytics
for Unstructured Data

TABLE OF CONTENTS



**PROJECT
INTRODUCTION**



**DATA
SOURCING**

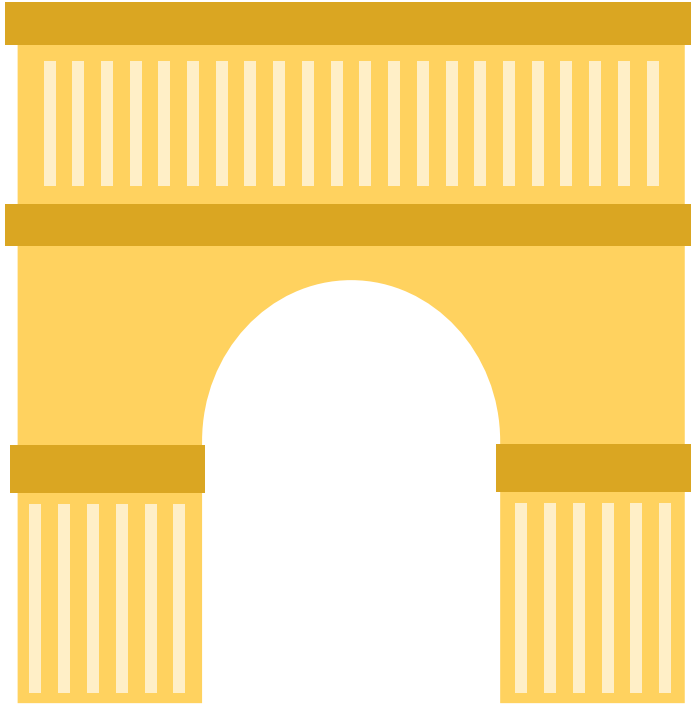


ANALYSIS



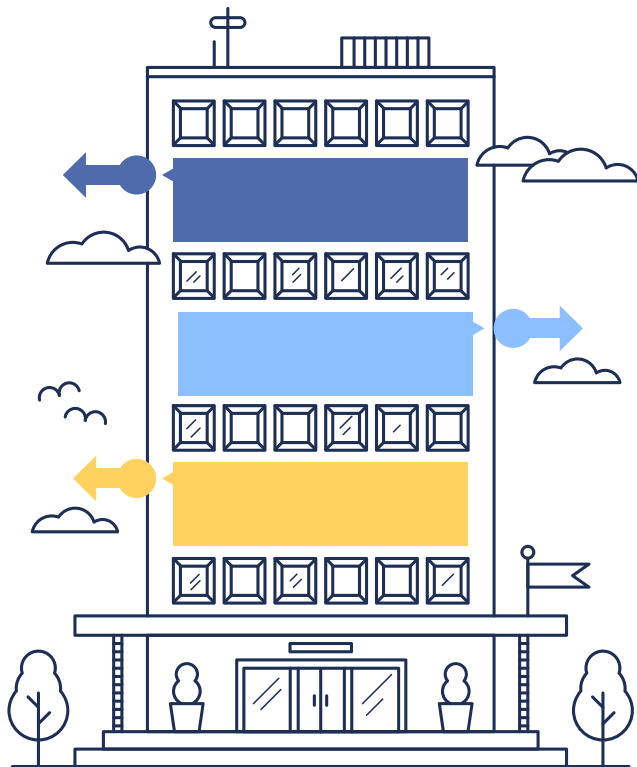
**FINAL
THOUGHTS**





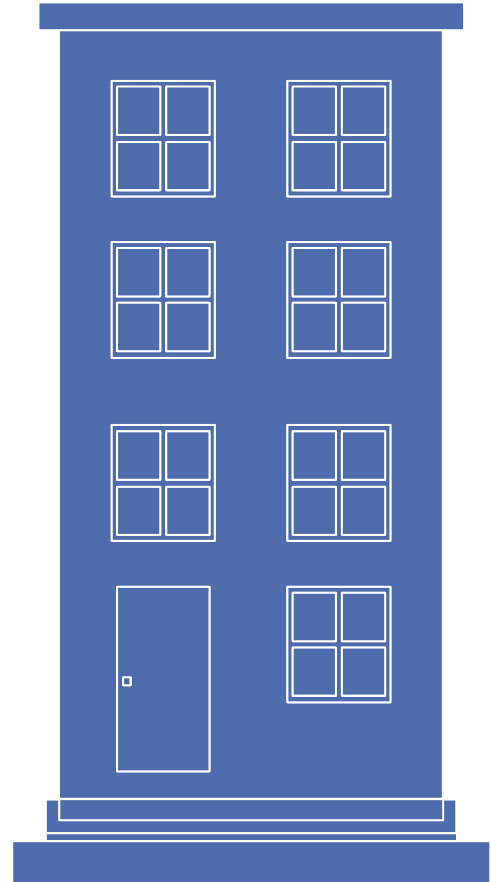
PROJECT INTRODUCTION

PROJECT INTRODUCTION



- Biggest question post-grad?
 - **Where should I live?**
- Some students stay in the same city and other move to experience a change
 - **Our Goal:** Help these students narrow down their choices based on how similar cities are and how the attributes they want fit into this

DATA SOURCING



GETTING OUR DATA

- Scraped data from various city subreddits
 - Used the current Top 10 cities, expanded to Top 20 for further analysis
 - Scraped information:
 - Title
 - Body
 - Up Votes
 - Top 5 Comments
- Based on scraped data....
 - word counts by city
 - total word counts
 - weighted the word counts for each based on upvotes
 - From this list identified key attributes



Sneak Peak

Scraped Posts/Comments

#	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Subreddit	Title	Body	Score	URL	Comment	Comment	Comment	Comment	Comment	Comment	Comment	Comment	Comment
2	seattle	Weekly As This three		1	https://ww	Hi all, lăC'm planning a first house purchase and am looking for any recommendations you have								
3	seattle	When you My wife cc		785	https://ww	I walked it I was tanr Today l wc My UV ligh- lăC's mid-October. Time for the long dark.								
4	seattle	Gold Creel My wife w		383	https://ww	Because a lt is so fru lăC'm sor The onlon Having worked for the national park service for seven)								
5	seattle	ALL MY WS *WANNA (424	https://ww	I watched My husbai Why not it When son lăC"ll have whatever drugs youăC're on. Please.								
6	seattle	WhatăC's is this a b.		50	https://ww	It's no mo 4:30am ar are you ta But l was i Belltown helicat now these guys. Sleepless in Seattle ii								
7	seattle	Rant! Seat But l donă		247	https://ww	So sorry tI see so m My rent (r If busines The Office of Economic Development (OED) Offers help								
8	seattle	Sub Sand Hello goo		7	https://ww	You hafta My Go-To I need to tI Tats is a g We are really lacking in that department. I like lyons d								
9	seattle	Monday C * As a note		7	https://www.reddit.com/r/Seattle/comments/1g3i6il/monday_capitol_hill_board_game_night_730_pm_at/									
10	seattle	Ballard ai Heard it la		20	https://ww	It's absol. I woke up Where is it? I'm not hearing it								
11	seattle	What is wI This is the		13	https://ww	Fire on N7 Came heri I love that https://we Police car choir practice								
12	seattle	Where to f I haven't dI		13	https://ww	RIP the my It's not qu Bowling a hospitals - ray gun lounge in capitol hill - Great wall mall has a								
13	seattle	What's yo I'm might		25	https://ww	Kenmore iI still have yo you hav lăC've bec DonăC" know if youăC've ever been to MoPOP, but ită								
14	seattle	Seattle Pa Has amoyr		1	https://www.reddit.com/r/Seattle/comments/1g3ktz2/seattle_parks_full_parking_lot_but_empty_park/									
15	seattle	Apologies You guys i		0	https://ww	How much! Yeah Oreg They're al: Given how I accept your apology and ask you to declare EVERY O'								
16	seattle	Miles "Hel At the time		210	https://ww	What will. Goddamn lăC'm sr Now there Fuck Miles Hudson. Fuckface.								
17	seattle	Where car ităC's all i		11	https://ww	Met Marke lăC"vs se Whole Fox Ballard Sa RIP the Hop-In								
18	seattle	Any self-ei trying t		18	https://ww	Did you cl lăC"m stu Not an ex Call them They just take your taxes so easily and make it extra cc								
19	seattle	Mental He Today i wI		607	https://ww	I'm so sor I had som Sorry you iBe sure yo I just recently finished a rehab stint because i just felt								
20	seattle	Ballard Br Construct		10	https://ww	Seeing the Fav bridge								
21	seattle	Local Roa I was in Cc		4	https://ww	Vita has tI My favorit Our local Olympia Ĥ You should check out Black Arrows Coffee! I bought sc								
22	seattle	PSA: Trافی Avoid 520.		912	https://ww	Not leavin Also bring I just got tI I just past Please tell people to avoid Capitol Hill too. lăC'm tytr								
23	seattle	I'm loving If people g		603	https://ww	The amazi I've been tI I'm loving as much I really love the dark and cold part of the year and all t								
24	seattle	Neighborl lăC've bec		8	https://cap Hill j. lăC've see Maple lea The Ballard burbs have great Halloween decorations. There is or									
25	seattle	I90 bridge Both direc		253	https://ww	OP, you ju Both direc DonăC" fr Both bridg Kayak time...								

Weighted Word Counts By City

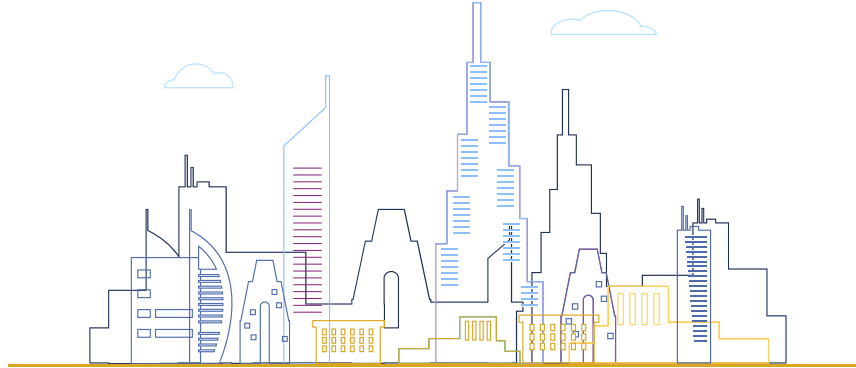
#	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Word	Count	Count	Count	Count	Count	Count	Count	Count	Count	Count	Count	Count	Count
2	the	234845	286720	37401	14903	95916	6217	10431	17429	45492	23235	56059	18525	26396
3	l	168650	197735	240875	88851	72910	38788	103685	116681	369280	133298	375299	232789	1101383
4	to	166203	203063	308787	87565	73605	33663	73257	136288	298687	140137	333592	113348	68082
5	and	153739	181195	139618	97824	64866	32533	77337	115076	340489	144775	335410	246916	132318
6	a	138900	165250	258606	84802	53952	36719	70079	124355	264992	124812	333298	254404	99548
7	of	94357	98422	173941	51598	38316	19764	33488	71643	207922	62056	234563	179703	79098
8	in	82600	96739	125756	53945	36479	17272	48470	82015	159994	74402	210635	135034	71443
9	it	78464	86786	133632	47467	40391	19663	45511	73089	185243	86413	202326	125403	69558
10	you	70306	72996	120731	54328	37191	17631	36136	59554	130133	55858	155995	112393	43857
11	is	64906	77194	95634	34648	28554	12791	19177	46619	133163	58772	165559	113375	49435
12	for	63753	75510	130363	41036	30844	14556	20308	50060	131502	54561	164296	91733	38785
13	so	57910	61287	108587	34474	22725	13012	28285	62398	113127	71072	164296	91733	38785
14	this	57259	63773	92252	33257	25779	10433	22949	56732	110913	61377	142865	73035	41383
15	that	52349	51266	89548	27845	24467	11359	25144	48556	83923	48632	140789	139082	34088
16	is	49474	44776	96883	30373	14128	7653	21463	37145	54405	167331	103233	30544	54294
17	t	44211	51435	67224	26148	20157	8509	21889	45979	65939	42336	94761	85668	32214
18	was	43881	44387	61645	30062	15365	12871	29920	35136	112906	30953	122836	77472	22148
19	they	41022	52591	80061	22967	18880	5966	18968	27052	58919	23560	94750	73404	29556
20	with	40680	40485	85445	20200	14904	7487	16899	24204	74176	28154	65486	61958	34077
21	my	39935	27117	87623	17623	19018	9135	13489	30027	84789	34025	121676	54111	29192
22	at	39933	40133	67282	21798	17902	13725	20263	24786	49911	32678	101814	54570	65027
23	but	38865	35123	57514	18108	16262	9064	15997	27610	70614	30202	91291	53895	21440
24	boston	34496	0	5	651	0	0	60	4	1558	0	2961	0	282

Word Counts By City

#	A	B	C	D	E	F	G	H	I	J	K	L	M	N
1	Seattle	Count	Count	Count	Count	Count	Count	Count	Count	Count	Count	Count	Count	Count
2	the	234845	286720	37401	14903	95916	6217	10431	17429	45492	23235	56059	18525	26396
3	l	1711	3832	2085	2404	4542	2658	358	4954	982	3558	3745	2388	2813
4	a	1448	3518	1932	2296	2786	2566	340	3074	990	3307	3410	2335	2573
5	i	1421	3547	1780	2116	2262	2619	231	3125	961	3475	3333	2539	2424
6	and	1365	3462	2099	2287	2585	2436	336	2990	967	3212	3456	1996	2477
7	of	891	1932	1275	1452	1792	1540	306	2187	528	1854	1931	1262	1302
8	in	890	2078	1077	1127	1244	1350	105	1789	628	1789	1742	1341	1399
9	it	870	2037	1517	1506	1656	1559	236	2421	576	1978	1518	1496	1582
10	you	773	2027	914	1066	1148	1282	134	1888	432	1654	1623	1116	1493
11	is	706	1481	840	1039	1066	1146	149	1528	400	1562	1404	981	1245
12	for	675	1778	936	1070	1266	1202	122	1066	411	1742	1780	1157	1374
13	so	561	1272	675	749	1858	811	129	1534	291	1073	1200	757	875
14	this	537	1266	720	904	900	876	99	873	368	1243	1204	888	788
15	that	531	1242	725	901	990	912	125	1075	367	1292	1150	954	931
16	t	471	962	488	567	864	652	53	730	231	885	950	678	672
17	but	409	982	498	613	594	680	78	749	279	968	924	693	721
18	this	396	878	457	543	1160	541	102	1384	281	731	864	497	354
19	have	360	1038	511	651	838	723	61	1235	273	1061	987	704	697
20	they	354	1024	517	607	574	611	69	633	293	901	914	633	508
21	if	353	896	406	594	760	614	99	1109	216	819	755	533	623
22	are	351	885	439	579	556	541	73	675	218	844	812	537	741
23	was	342	754	387	446	706	514	55	895	225	689	623	515	442
24	with	339	859	529	587	498	577	89	566	255	812	993	514	575
25	or	337	860	1092	496	512	567	63	966	248	741	844	532	962
26	at	335	818	419	508	588	570	48	1473	217	780	693	433	631
27	he	323	624	441	574	1096	560	58	1007	201	727	715	692	543
28	my	313	914	160	513	489	660	80	597	353	797	567	567	492
29	can	295	683	297	399	354	476	66	500	162	621	605	371	482
30	seattle	290	0	0	1	2	0	0	1	1	1	0	0	1

LIST OF TOP CITIES

AUSTIN
DALLAS
HOUSTON
CHICAGO
SAN DIEGO
LOS ANGELES
PHOENIX
SEATTLE
PHILADELPHIA
NEW YORK CITY



BOSTON
PORTLAND
(OREGON)
DENVER
NASHVILLE
LAS VEGAS
DETROIT
COLUMBUS
JACKSONVILLE
SAN FRANCISCO

ATTRIBUTES/FEATURES



SAFE



HOUSING



HOT



JOBS



RESTAURANTS



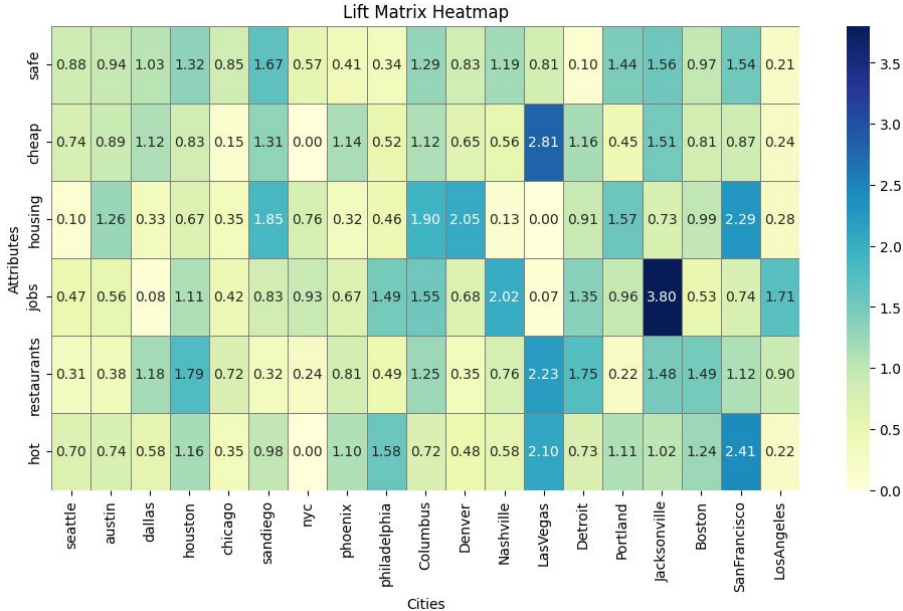
CHEAP



OUR ANALYSIS

Lift Matrices

Lift Matrix:																			
	seattle	austin	dallas	houston	chicago	sandiego	nyc	phoenix	philadelphia	Columbus	Denver	Nashville	LasVegas	Detroit	Portland	Jacksonville	Boston	SanFrancisco	LosAngeles
safe	0.875485	0.943478	1.034941	1.323834	0.845319	1.667589	0.569676	0.414578	0.342942	1.291498	0.831891	1.191897	0.809546	0.097815	1.438962	1.557427	0.971805	1.540727	0.210434
cheap	0.744673	0.891674	1.117846	0.828345	0.147490	1.313355	0.000000	1.141861	0.518579	1.117799	0.653162	0.563226	2.805350	1.164801	0.445075	1.506546	0.808232	0.873678	0.238656
housing	0.097193	1.256894	0.328272	0.668952	0.346501	1.851291	0.758918	0.315599	0.456865	1.901631	2.045978	0.132320	0.000000	0.912161	1.568433	0.732280	0.992550	2.294020	0.280339
jobs	0.474984	0.563060	0.080214	1.114496	0.423340	0.829337	0.927214	0.674774	1.488474	1.548888	0.676999	2.020780	0.073202	1.353249	0.958122	3.802338	0.527241	0.737562	1.712528
restaurants	0.311188	0.375598	1.177170	1.791341	0.721117	0.316127	0.242987	0.808376	0.487590	1.246704	0.354831	0.762579	2.225264	1.752310	0.223188	1.484900	1.492232	1.121062	0.897575
hot	0.697876	0.735363	0.576179	1.164435	0.345554	0.984657	0.000000	1.101578	1.579471	0.722452	0.476091	0.580615	2.103246	0.727735	1.112283	1.022390	1.239449	2.408160	0.223658



Recommendation Model



For the cost-conscious restaurant goer (attributes are “bars”, “cheap”, “restaurants”)

Recommendations Based on Your Preferences:

City	Similarity	Average_Sentiment_Score	Overall_Score
Jacksonville	0.392390	0.200902	0.078832
LasVegas	0.421752	0.181574	0.076579
sandiego	0.386492	0.194135	0.075032
Columbus	0.401450	0.177279	0.071168

For the grad student looking to settle down (attributes are “clean”, “safe”, “kids”)

Recommendations Based on Your Preferences:

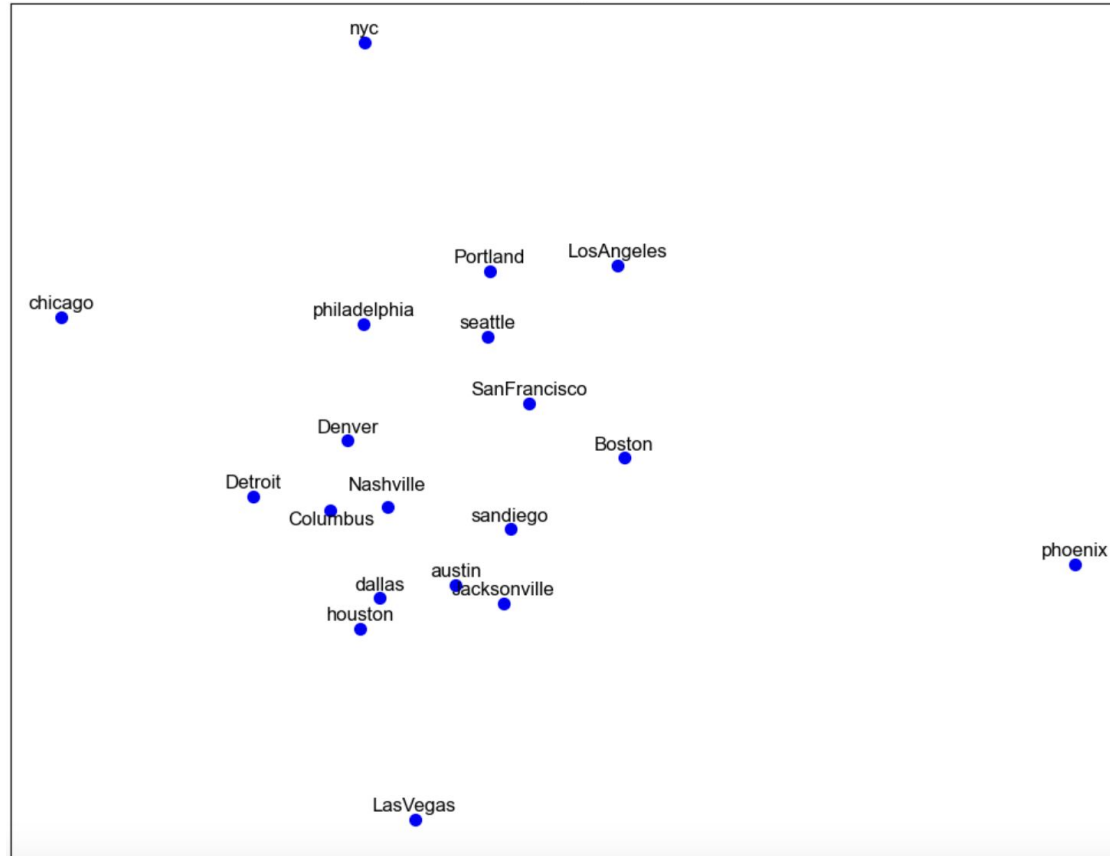
City	Similarity	Average_Sentiment_Score	Overall_Score
phoenix	0.195308	0.176652	0.034502
Portland	0.201568	0.136471	0.027508
seattle	0.160669	0.158653	0.025491

For the active socializer (attributes are “sports”, “theatre”, “bars”)

Recommendations Based on Your Preferences:

City	Similarity	Average_Sentiment_Score	Overall_Score
LasVegas	0.269316	0.181404	0.048855
SanFrancisco	0.241389	0.185089	0.044679
Detroit	0.231839	0.180810	0.041919
chicago	0.226385	0.164338	0.037204

MDS Plot of City Similarities



Pairing Systems

User-Input Driven

Enter a city name: houston

Top 3 Big Cities closest to 'houston':

dallas
austin
sandiego

Top 3 Small Cities closest to 'houston':

Columbus
Nashville
Jacksonville



Static

City Pairings Based on MDS Similarities:

Big City	Paired Big City	Paired Small City
Losangeles	Seattle	Portland
Austin	Dallas	Jacksonville
Chicago	Philadelphia	Detroit
Dallas	Houston	Nashville
Houston	Dallas	Columbus
New York	Philadelphia	Portland
Philadelphia	Seattle	Denver
Phoenix	Losangeles	Boston
San Diego	Austin	Jacksonville
Seattle	Philadelphia	Portland

City Pairings Based on MDS Similarities:

Small City	Paired Small City	Paired Big City
Columbus	Nashville	Dallas
Denver	Columbus	Philadelphia
Nashville	Columbus	Dallas
Lasvegas	Jacksonville	Houston
Detroit	Columbus	Dallas
Portland	Sanfrancisco	Seattle
Jacksonville	Nashville	Austin
Boston	Sanfrancisco	San Diego
Sanfrancisco	Boston	Seattle

FINAL THOUGHTS



THANK YOU



FOR LISTENING