**APAN 5310
Project Report**

NYC DOB Construction Job Filings Database

Kexin lian, Dongquan Qiu, Wenqian Liu, Junyao Xie, Chunye Xie

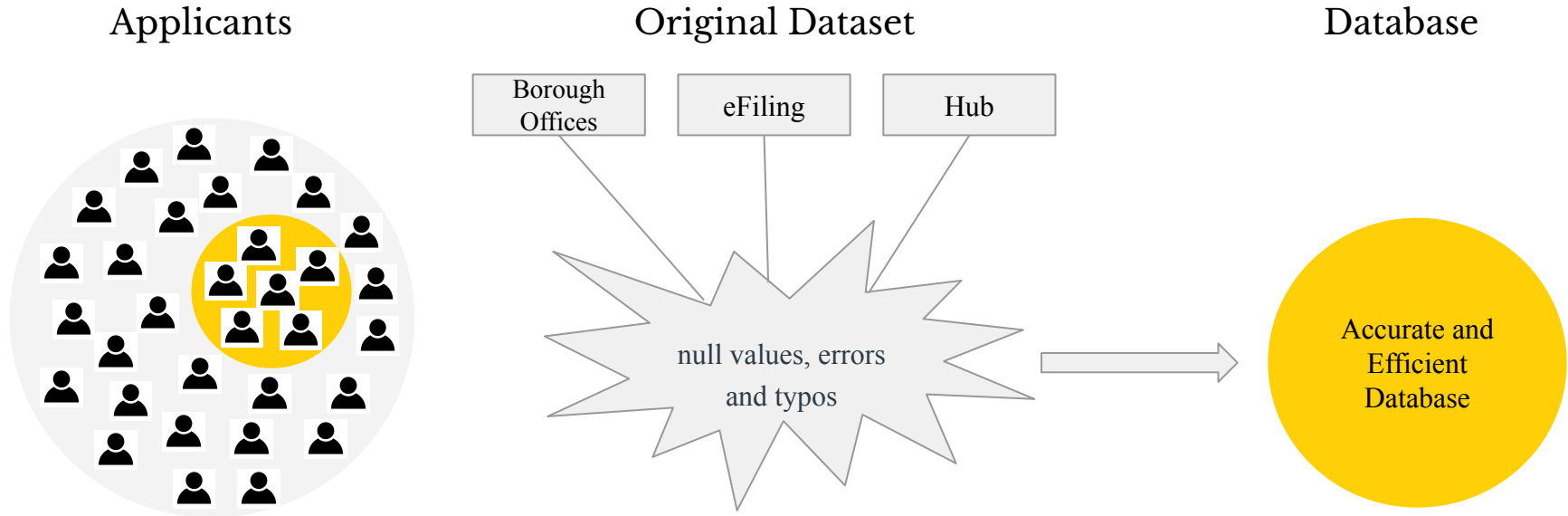Aug 10, 2020

Agenda

- ❏ Scenario
- ❏ Original Dataset Sample
- ❏ Normalization Plan & ER Diagram
- ❏ ETL Process
- ❏ Analytics Procedures
- ❏ Dashboard Demo

Scenario - NYC DOB Construction Job Filings

Applicants

Original Dataset

Database

Borough Offices

eFiling

Hub

null values, errors and typos

Accurate and Efficient Database

- Only approve a few applicants from the large applicant pool
- We are hired to create an accurate and efficient database

Original Dataset Sample

doc	fee_status	job_id	job_status	latest_action_date	professional_cert	prefiling_date
01	STANDARD	110083524	X	4/8/09 0:00	Y	2/4/08
01	STANDARD	103039766	X	12/19/02 0:00	Y	12/10/01
01	EXEMPT	102616885	R	5/3/00 0:00	NaN	5/2/00

paid_date	fully_paid_date	borough_name	house_number	building_bin	job_street_name	owner_company1_name	owner_company2_name
2/4/08	2/4/08	MANHATTAN	152	1084455	WEST 57 STREET	Carnegie Hall Towers	LYCEUM THEATER CORP
12/10/01	12/10/01	MANHATTAN	280	1035441	PARK AVENUE	Boston Properties	
5/2/00	5/2/00	MANHATTAN	600	1019483	THIRD AVENUE	Gale & Wen	AMERICAN MUSEUM OF NATURAL
5/2/00	5/2/00	BROOKLYN	22	4003540	30 PLACE	30th Place Holdings LLC	OCEAN AVENUE LLC

owner_first_name	owner_last_name	street_name	city	state	zip	building_type	landmarked	permit_type	unemployment_rate
Michael	Taub	152 West 57th Street	New York	NY	10019	OTHERS	N	EW	16
Michael	Taub	152 West 57th Street	New York	NY	10019	OTHERS	N	EW	16

labor_force	owner_cell_phone	owner_home_phone	permittee_first_name	permittee_last_name	permittee_phone	permit_status	permit_subtype	permit_sequence
848.9	2126031610	2128793672	STEPHEN	EISNER	2126974422	ISSUED	OT	1
848.9	2123086661	2128881234	CHI	CHAN	6463019922	ISSUED	OT	2
848.9	2129867787	2120372937	ROBERT	SCHUBERT	5162421188	ISSUED	FP	1

filing_rep_first_name	filing_rep_last_name	manager_first_name	manager_last_name	applicant_first_name	applicant_last_name	applicant_company_name	applicant_professional_title	applicant_license_number
WILLIAM	VITACCO	Alissa	Morrow	Bruce	Lilker	Build it brother	PE	60859
BALDO	SACCHERI	Hibba	Swanson	MATT	MARKOWITZ	inspiration design build	RA	22409
NaN	NaN	Giorgia	Rubio	Frank	Eilam	blue ladder construction	RA	24701
WILLIAM	VITACCO	Tayler	Stott	Ravi	Shenoy	builder gorilla	PE	55232

gis_latitude	gis_longitude	manager_company_name	owner_company1_type	owner_company2_type	work_type1	work_type2
40.68725	-73.974168	Victoria Restoration Corp	CORPORATION	INDIVIDUAL	OT	
40.747521	-73.985239	50 SUTTON PLACE SO. OWNER	PARTNERSHIP	CORPORATION	OT	
40.751281	-73.981593	C/O BLDG Management Co., Inc	INDIVIDUAL	CORPORATION	FP	PL
40.792464	-73.964113	The Estate of Vincent Terranva	CORPORATION	PARTNERSHIP	EQ	BL
40.840096	-73.915569	ASTERDAM PEDIATRIC	CORPORATION	PARTNERSHIP	OT	SP

(63244)

- Original: 50 columns & 1 million rows
- We only insert 10% of the data into the database.

1NF

manager_company_name	owner_company1_type	owner_company2_type	w
Victoria Restoration Corp	CORPORATION	INDIVIDUAL	
50 SUTTON PLACE SO. OWNER	PARTNERSHIP	CORPORATION	
C/O BLDG Management Co., Inc.	INDIVIDUAL	CORPORATION	
The Estate of Vincent Terranva	CORPORATION	PARTNERSHIP	
ASTERDAM PEDIATRIC	CORPORATION	PARTNERSHIP	

building_bin	job_street_name	owner_company1_name	owner_company2_name	owner_first_name
1084455	WEST 57 STREET	Carnegie Hall Towers	LYCEUM THEATER CORP	Michael
1035441	PARK AVENUE	Boston Properties		Michael
1019483	THIRD AVENUE	Gale & Wen	AMERICAN MUSEUM OF NATURA	Robert
4003540	30 PLACE	30th Place Holdings LLC		Mark
1012268	WEST 26 STREET	RXR SL OWNER LLC	DK 1191 OCEAN AVENUE LLC	MITCHELL

2NF

Initial table

doc	job_id	job_status	latest_action_date	fully_paid_date	borough_name	house_number	building_bin	job_street_name
01	110083524	X	4/8/09 0:00	2/4/08	MANHATTAN	152	1084455	WEST 57 STREET
01	103039766	X	12/19/02 0:00	12/10/01	MANHATTAN	280	1035441	PARK AVENUE
01	102616885	R	5/3/00 0:00	5/2/00	MANHATTAN	600	1019483	THIRD AVENUE
01	102616894	P	8/29/01 0:00	5/2/00	BROOKLYN	22	4003540	30 PLACE
01	402461180	P	9/25/06 0:00	9/20/06	STATEN ISLAND	601	1012268	WEST 26 STREET

Job_info in 2NF

job_id	doc	fee_status	job_status	latest_action_date	professional_cert
110083524	01	STANDARD	X	4/8/09 0:00	Y
103039766	01	STANDARD	X	12/19/02 0:00	Y
102616885	01	EXEMPT	R	5/3/00 0:00	NaN
102616894	01	EXEMPT	P	8/29/01 0:00	NaN
402461180	01	EXEMPT	P	9/25/06 0:00	NaN

Job location in 2NF

job_id	borough_name	unemployment_rate	labor_force	building_bin	job_street_name
110083524	MANHATTAN	4.4	848.9	1084455	WEST 57 STREET
103039766	MANHATTAN	4.4	848.9	1035441	PARK AVENUE
102616885	MANHATTAN	4.4	848.9	1019483	THIRD AVENUE
102616894	BROOKLYN	6.6	1171.5	4003540	30 PLACE
402461180	STATEN ISLAND	6	211	1012268	WEST 26 STREET

3NF

Building in 2NF

building_bin	building_type	landmarked	permittee_first_name	permittee_last_name	permittee_phone
1084455	OTHERS	N	STEPHEN	EISNER	2126974422
1035441	OTHERS	N	CHI	CHAN	6463019922
1019483	OTHERS	NaN	ROBERT	SCHUBERT	5162421188
4003540	OTHERS	N	MICHAEL	PALADINO	6318421700
1012268	OTHERS	NaN	SANG	KIM	7184295000

Permittee in 3NF

permittee_id	permittee_first_name	permittee_last_name	permittee_phone
1	STEPHEN	EISNER	2126974422
2	CHI	CHAN	6463019922
3	ROBERT	SCHUBERT	5162421188
4	MICHAEL	PALADINO	6318421700
5	SANG	KIM	7184295000

ER Diagram
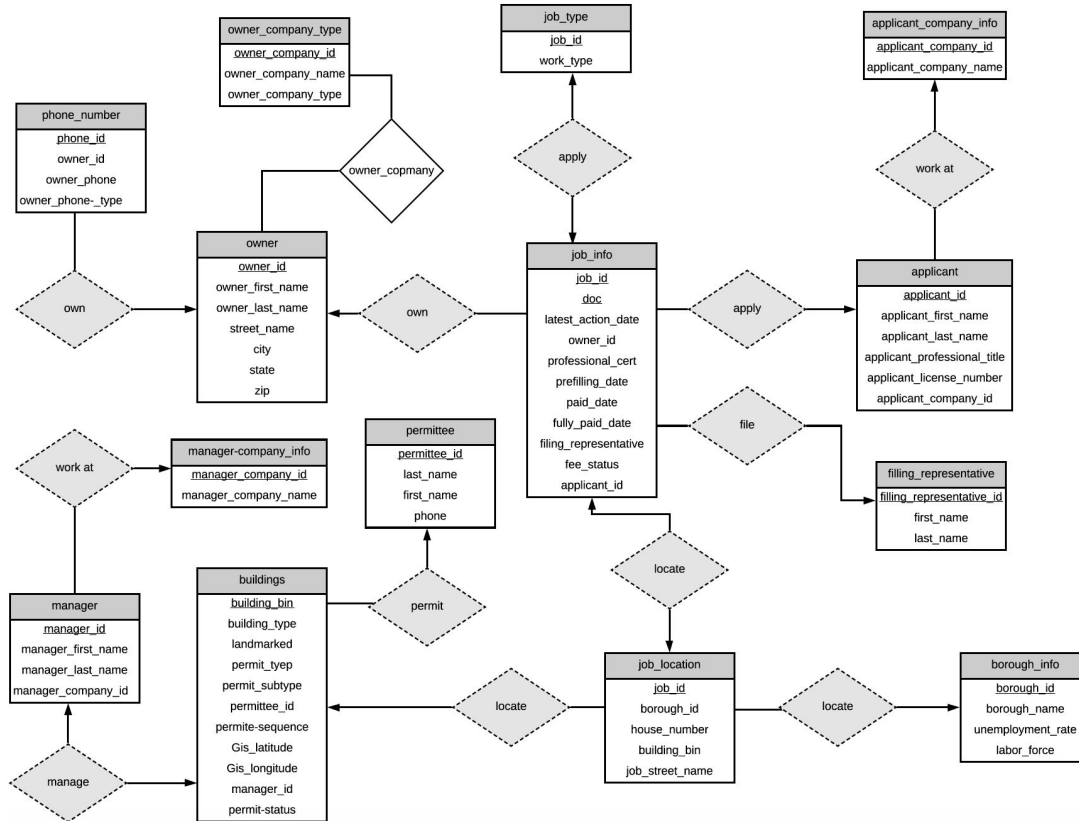
ETL Process

✓ Import the necessary packages

```
import pandas as pd
import numpy as np

from sqlalchemy import create_engine
```

✓ Extract, Transform and Load

- Explore the data to ensure proper table design
- Extract the data based on table design
- Process and transform the data
  - Add primary key
  - Data type modification
  - Error correction
- Load data into PostgreSql database

```
applicant_company_df.to_sql(name='applicant_company_info', con=engine, if_exists='append', index=False)
```

✓ Connect with the database

```
conn_url = 'postgresql://postgres:123@localhost:5432/Sql_Project'

engine = create_engine(conn_url)
connection = engine.connect()
```

Sample ETL Process

applicant_company_info

```
create table applicant_company_info(
    applicant_company_id integer,
    applicant_company_name varchar(200),
    PRIMARY KEY (applicant_company_id)
);
```

```
[4]: # Select relevant columns and add id
applicant_company_df = df[['applicant_company_name']].drop_duplicates().dropna()
applicant_company_df.insert(0, 'applicant_company_id', range(1, 1 + len(applicant_company_df)))
```

```
[5]: applicant_company_df
```

[5]:	applicant_company_id	applicant_company_name
0	1	Anthony Rubbish Removal
1	2	Walsh Construction
2	3	Hockman Hill Group
3	4	S&Z Construction
4	5	Mectec NYC, Inc.

Analytics Procedures

Job Application:

1. How many jobs are professionally certified? 51046
2. What is the most common work type? OT, PL, EQ, MH, and SP
3. How many jobs are available in the last 5 years? Increasing (~225% compounded growth)
4. How many jobs available are there in each borough in NYC? Queen most, Manhattan

Building:

5. How many residential buildings (family) are in job applications? 1-2-3 FAMILY(~44.9%)
6. What are the most popular streets for building construction? BROADWAY street
7. Top 5 owner companies that own the most constructions? NY School Construction Authority, NYC HPD
8. Top 5 manager companies that engaged the most construction? BEECH ASSO
9. How does the top 1 company distribute its construction on Map? MAJORITY:Manhattan and Queens
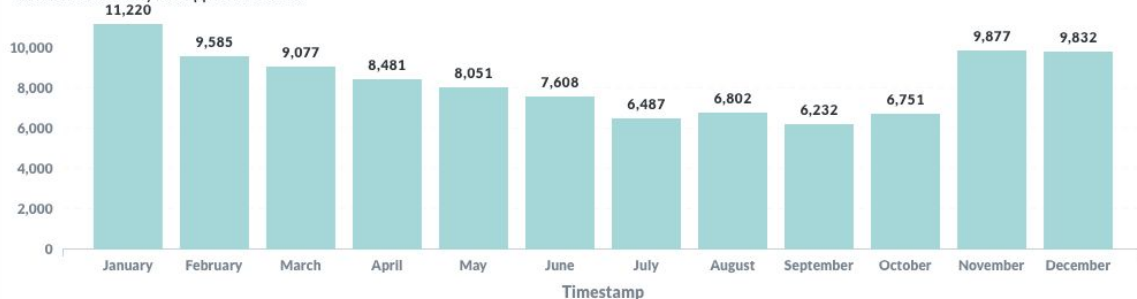10. What is the trending of job markets and which area can be the most potential one?
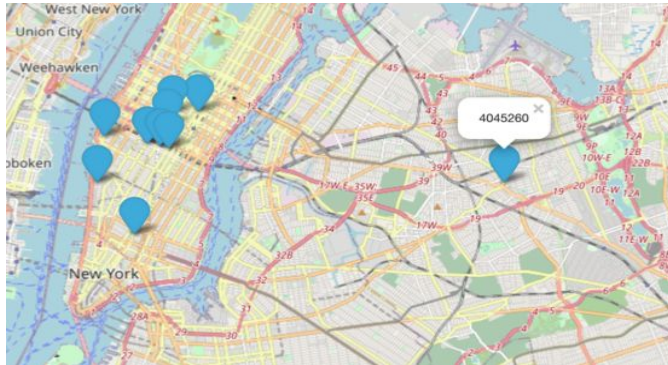
- **BEECH ASSO engaged the most construction buildings work as manager companies.**
- **The distribution of BEECH ASSO are mainly in Manhattan and Queens area.**
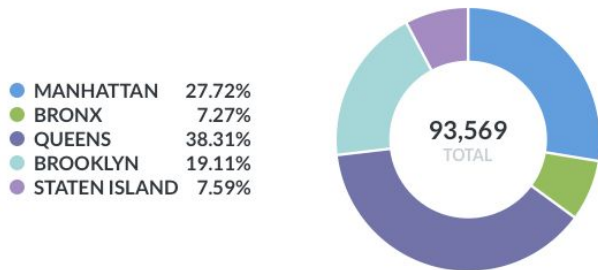
Dashboard Demo

5310_Project
≋ Our analytics

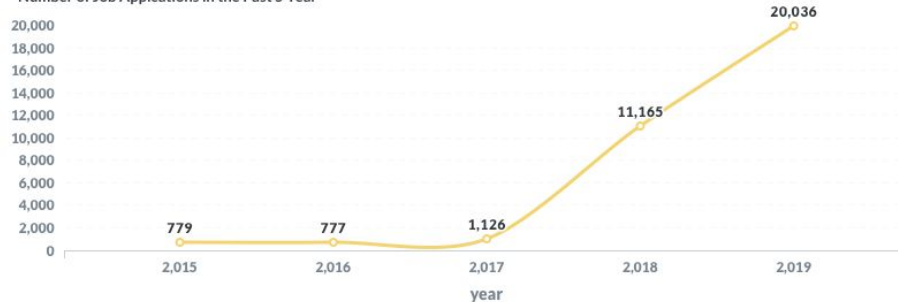### Cumulative Monthly Job Application Trend

Month	Value
January	11,220
February	9,585
March	9,077
April	8,481
May	8,051
June	7,608
July	6,487
August	6,802
September	6,232
October	6,751
November	9,877
December	9,832

Timestamp

Top 5 Manager Companies that Manage the Most Buildings

- BEECH ASSO
- 4 STAR CONTRACTING INC
- 2CONTRUCTION REALITY SAFETY GROUP
- ADM CONSTRUCTION CO. LLC
- CR SAFETY GROUP INC

Job Application Number per Borough in NYC

- ● MANHATTAN   27.72%
- ● BRONX   7.27%
- ● QUEENS   38.31%
- ● BROOKLYN   19.11%
- ● STATEN ISLAND   7.59%

93,569
TOTAL

Number of Job Appications in the Past 5 Year

year	Value
2,015	779
2,016	777
2,017	1,126
2,018	11,165
2,019	20,036

We appreciate your feedbacks and are open to questions...

THANKS