where(ii)

By Melanie, Lisa, Priya

+

+ $ +

Major Metro Regions

Rent Prices

Income

Traffic

The where ii project aims to help people improve their quality of life by identifying locations in which they can maximize their income relative to the cost of rent, factoring in time spent in traffic and rental trends over time.

where(ii) Milestones

October

- Finalized data sources
- Utilized Python/Pandas to scale data cleansing
- Created ERD
- Created DDL scripts

November

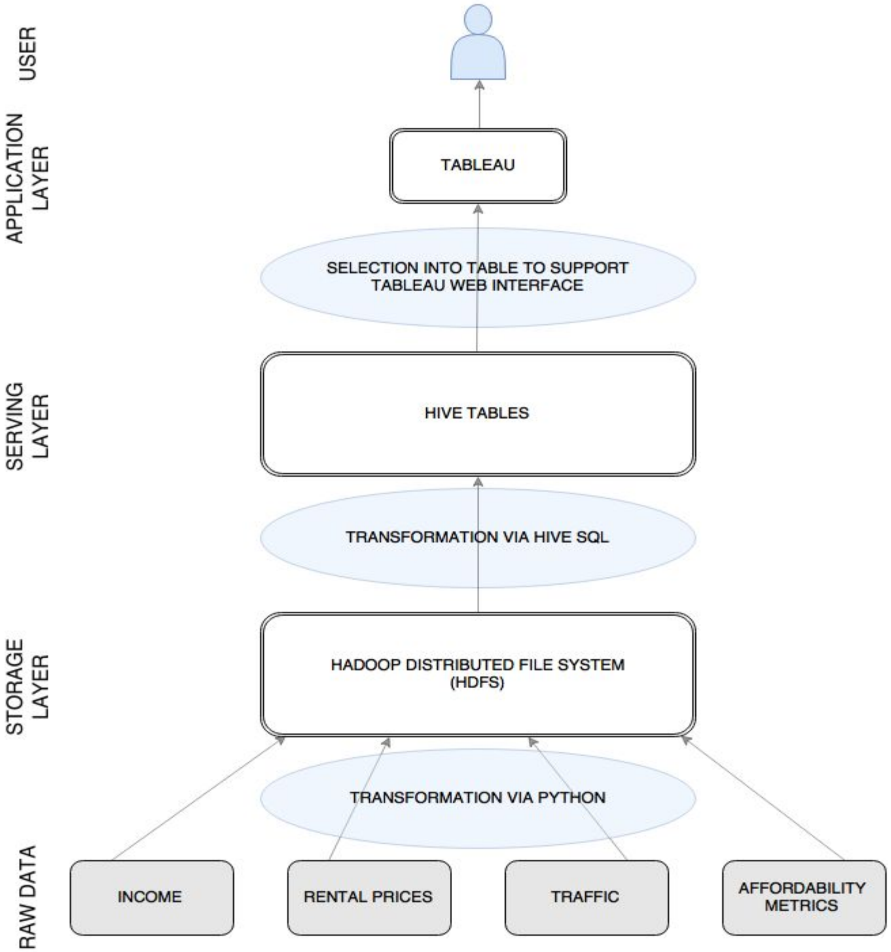
- Tested .sh scripts
- Updated ERD
- Partition tables
- Finalize ETL
- Created Tableau Dashboard

December

- Updated ERD
- Publish Tableau Dashboard
- Use Case Testing
- Finalized .sh scripts

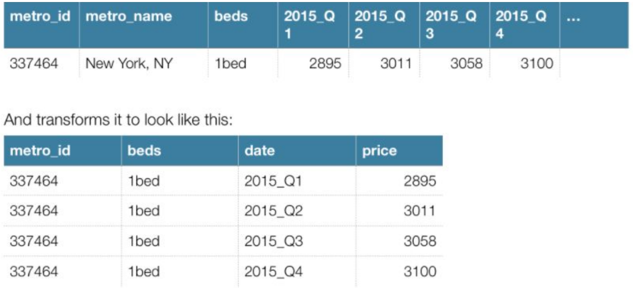
Storage and Retrieval

- Tools Utilized:
  - HDFS
  - Hive
  - Tableau
  - Python/pandas
  - SQL
- RDBMS approach
  - Data contains unique identifiers
  - Raw data sources are topic specific
- Tableau
  - Interactive way to explore and engage
  - User can quickly scale through different occupations, housing types, and even salary ranges


USER
APPLICATION LAYER
TABLEAU
SELECTION INTO TABLE TO SUPPORT TABLEAU WEB INTERFACE
SERVING LAYER
HIVE TABLES
TRANSFORMATION VIA HIVE SQL
STORAGE LAYER
HADOOP DISTRIBUTED FILE SYSTEM (HDFS)
TRANSFORMATION VIA PYTHON
RAW DATA
INCOME
RENTAL PRICES
TRAFFIC
AFFORDABILITY METRICS

Database Transformations

Cleaning via Python



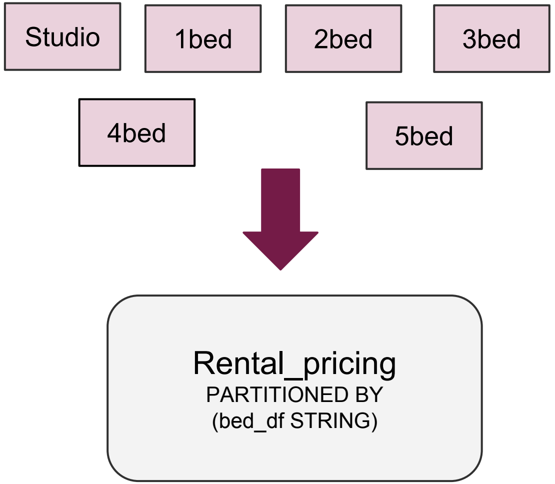
metro_id	metro_name	beds	2015_Q1	2015_Q2	2015_Q3	2015_Q4	...
337464	New York, NY	1bed	2895	3011	3058	3100	

And transforms it to look like this:

metro_id	beds	date	price
337464	1bed	2015_Q1	2895
337464	1bed	2015_Q2	3011
337464	1bed	2015_Q3	3058
337464	1bed	2015_Q4	3100

- ➔ Utilizing inputs at the command line via argparse
- ➔ Pandas melt function
- ➔ Scalable for all future Zillow files
- ➔ Created several .py files:
  - ◆ Zillow_process.py
  - ◆ Clean_mapping.py
  - ◆ Traffic_clean_transform.py

Partitioning



Studio

1bed

2bed

3bed

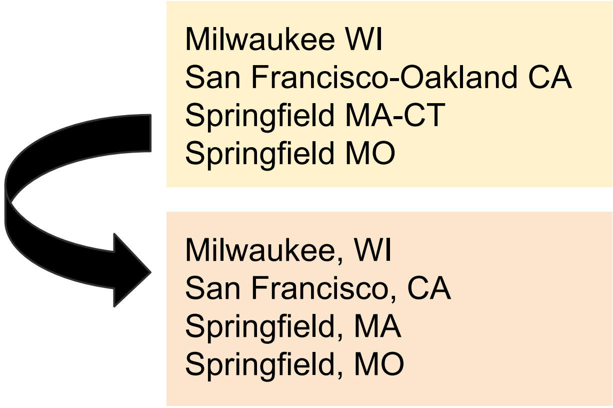
4bed

5bed

Rental_pricing
PARTITIONED BY
(bed_df STRING)

- ➔ Compiled all housing types into a single table
- ➔ Partitioned on type
- ➔ Created in Hive ETL

String Matching



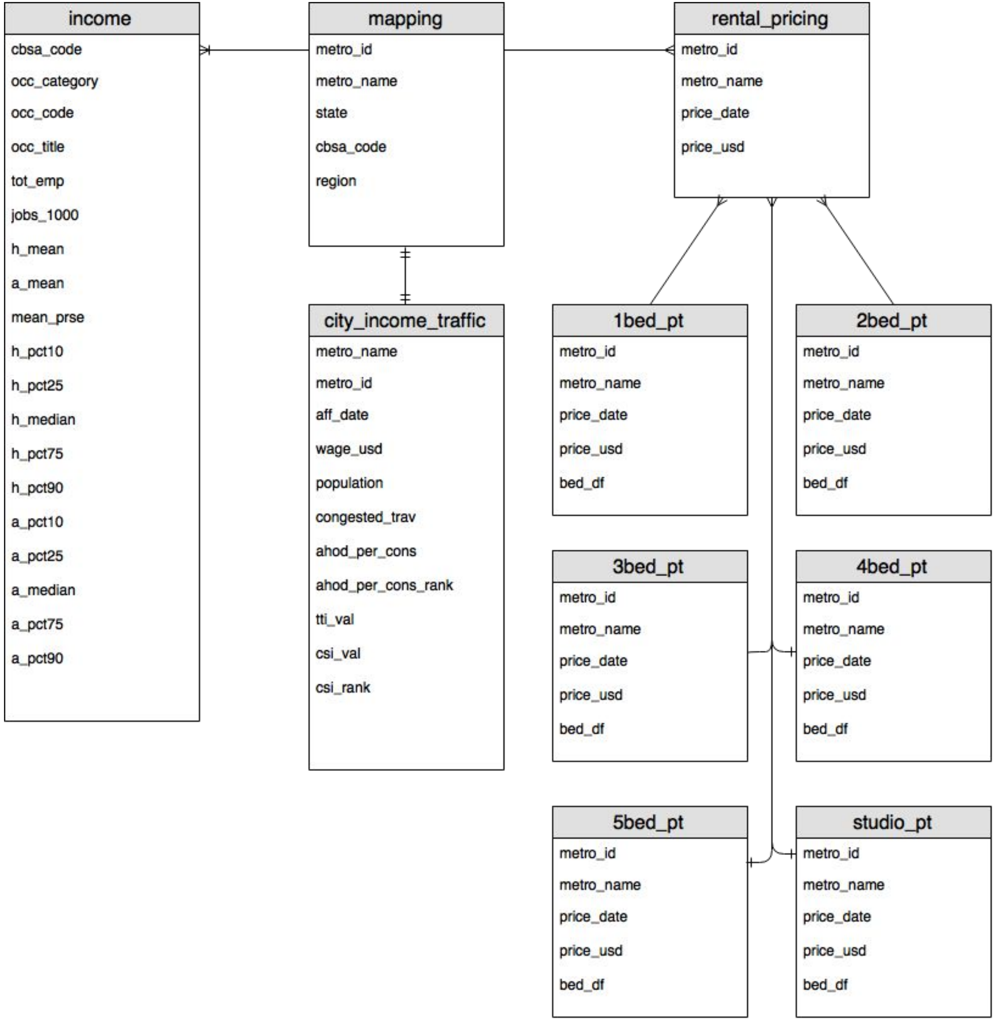
Milwaukee WI
San Francisco-Oakland CA
Springfield MA-CT
Springfield MO

Milwaukee, WI
San Francisco, CA
Springfield, MA
Springfield, MO

- ➔ Identify different metro regions amongst data sources
- ➔ Used Python difflib.sequencematcher
- ➔ Matched against zillow income data to find best match

ERD

- mapping table connects sources with different identifiers
- 1bed, 2bed, etc. feed into partitioned rental_pricing table
- Derived values in mapping table (region) and income table (occ_category) allow for easy filtering



income
cbsa_code
occ_category
occ_code
occ_title
tot_emp
jobs_1000
h_mean
a_mean
mean_prse
h_pct10
h_pct25
h_median
h_pct75
h_pct90
a_pct10
a_pct25
a_median
a_pct75
a_pct90

mapping
metro_id
metro_name
state
cbsa_code
region

rental_pricing
metro_id
metro_name
price_date
price_usd

city_income_traffic
metro_name
metro_id
aff_date
wage_usd
population
congested_trav
ahod_per_cons
ahod_per_cons_rank
tti_val
csi_val
csi_rank

1bed_pt
metro_id
metro_name
price_date
price_usd
bed_df

2bed_pt
metro_id
metro_name
price_date
price_usd
bed_df

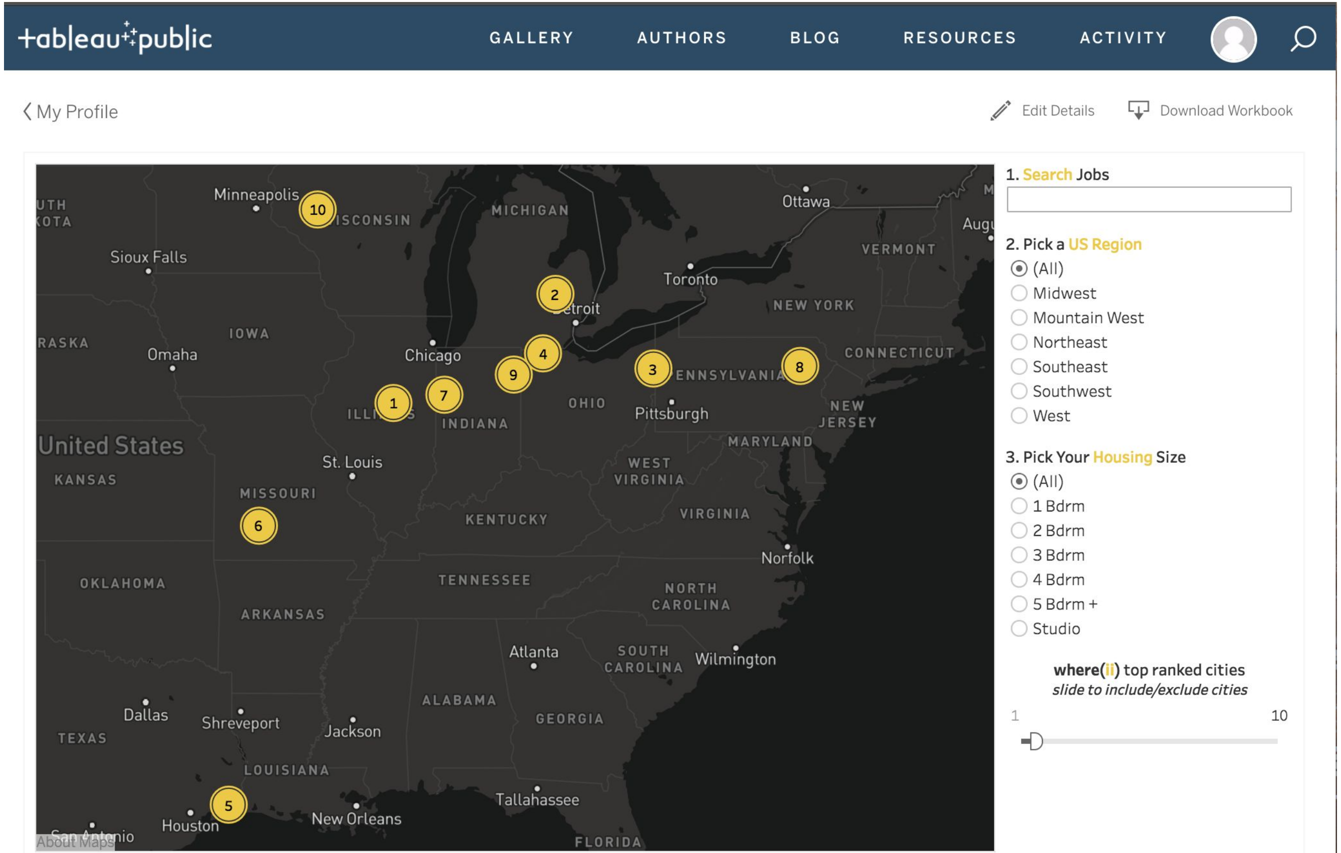
3bed_pt
metro_id
metro_name
price_date
price_usd
bed_df

4bed_pt
metro_id
metro_name
price_date
price_usd
bed_df

5bed_pt
metro_id
metro_name
price_date
price_usd
bed_df

studio_pt
metro_id
metro_name
price_date
price_usd
bed_df

The Final Product



1. Search Jobs

□

2. Pick a US Region

- ● (All)
- ○ Midwest
- ○ Mountain West
- ○ Northeast
- ○ Southeast
- ○ Southwest
- ○ West

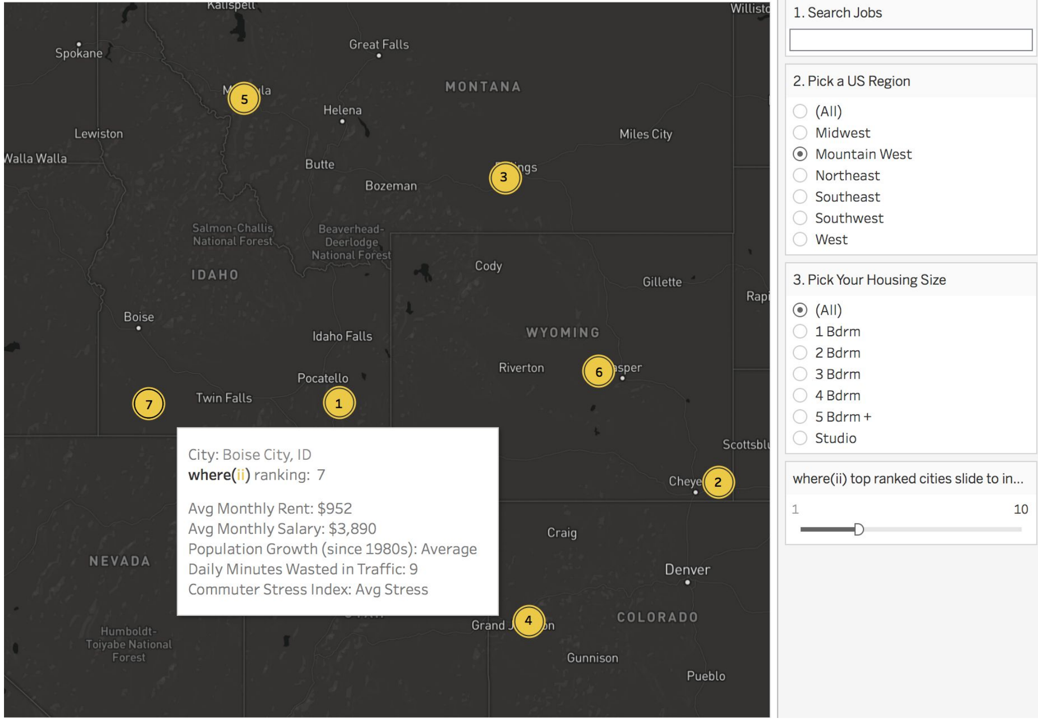
3. Pick Your Housing Size

- ● (All)
- ○ 1 Bdrm
- ○ 2 Bdrm
- ○ 3 Bdrm
- ○ 4 Bdrm
- ○ 5 Bdrm +
- ○ Studio

where(ii) top ranked cities
slide to include/exclude cities

1 10

where(ii)



1. Search Jobs

2. Pick a US Region

- ○ (All)
- ○ Midwest
- ● Mountain West
- ○ Northeast
- ○ Southeast
- ○ Southwest
- ○ West

3. Pick Your Housing Size

- ● (All)
- ○ 1 Bdrm
- ○ 2 Bdrm
- ○ 3 Bdrm
- ○ 4 Bdrm
- ○ 5 Bdrm +
- ○ Studio

where(ii) top ranked cities slide to in...

1 10

City: Boise City, ID
where(ii) ranking: 7

Avg Monthly Rent: $952
Avg Monthly Salary: $3,890
Population Growth (since 1980s): Average
Daily Minutes Wasted in Traffic: 9
Commuter Stress Index: Avg Stress

Tableau Public Link

https://public.tableau.com/profile/publish/205ProjWkbk_Mini/whereii#!/