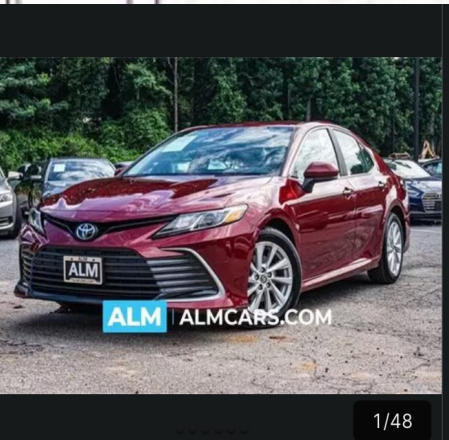
Apartments for Sale-Real Estate Market in Riyadh

Introduction

1) Provide home-buyers with tools that help them better understand the target market (apartments for sale in Riyadh) and **2)**, get an estimate of how fair the asking price is for an apartment



Used
Save
2021 Toyota Camry LE
50,393 mi.
$24,977 **$1,000 price drop**
$331 est./mo.*
Great Deal | $969 under
Home Delivery
Virtual Appointments
Free CARFAX 1-Owner Report
ALM Marietta
4.7 (3,026 reviews)
14 mi. from 30322
Check availability

Goal: Develop an analytical dashboard and create a “Good Deal Indicator” feature.

Community Engagement: Contribute a clean dataset on Kaggle

1 Data Dictionary

2 Data Web Scraping

3 Data Cleaning

4 Machine Learning model

5 “Good Deal Indicator”

6 Analytical Dashboard

7 CONCLUSION

OUTLINES

DATASET

Features	Description
user_id	The ad poster's Aqar ID
id	A unique value for each ad
price	Sale value in SR
beds	Number of bedrooms
livings	Number of living rooms
wc	Number of bathrooms
area	In sq. meters
street_width	Of the apartment complex, in meters
age	In years
street_direction	Of the apartment complex, (e.g., north-south, east-west)

DATASET

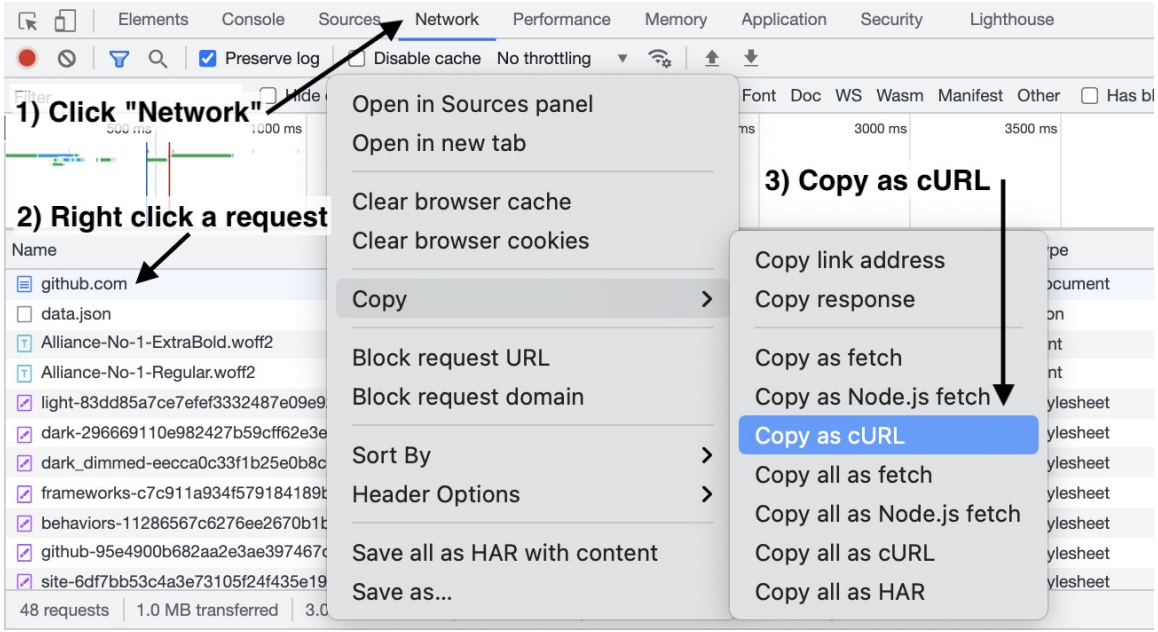
Features	Description
ketchen	{0, 1}
furnished	{0, 1}
district	The name of the district in Arabic
width	In meters
length	In meters
advertiser_type	{‘owner’, ‘exclusive_advertiser’, etc}
city_side	{‘north’, ‘south’, etc}
district_en	The name of the district in English
time_on_market	Number of days since post date until today
time_since_update	Number of days since the post was last updated

DATASET

Features	Description
last_update	Four features: year, month, day, hour
create_time	Four features: year, month, day, hour
latitude	Latitude coordinate of apartment complex
longitude	Longitde coordinate of apartment complex
squareness	Does the apartment look more like a square or rectangle? The ratio of length to width
regular_shapeness	Is the apartment shape regular? The ratio of area to length * width

Data Web Scraping

1) Identify the relevant request



Data Web Scraping

2) Make a python request and save to csv file

```
response = requests.post('https://sa.aqar.fm/graphql', cookies=cookies, headers=headers, json=json_data)
clean = False

# a webpage containing 20 apartment listings and other elements
response_json = response.json()
# getting relevant information
listings_list = response_json.get('data').get('Web').get('find').get('listings')

listings_list = pd.DataFrame(listings_list)

# appending listings from different pages to one dataframe
if i != 0:
    if list(all_listings.iloc[0])[1] ==list(listings_list.iloc[0])[1]:
        print("here is the problem")
        mesege = i
        break

all_listings = all_listings.append(listings_list, ignore_index=True)
if desired_post_id in list(listings_list['id']):
    return listings_list.loc[listings_list['id'] == desired_post_id], clean
elif i >= 500:
    return "Search failed", clean

sleep(0.5)
i+=20
```

Data Cleaning

- Drop Irrelevent features (e.g., desription, images)
- Find the district’s name in English (using outside sources)
- Create time features (change from integer to multiple informative features)
- Extract coordinates (from a string)
- Create experimental features (e.g., squareness)

Data Assumptions

- Advertisers provide fair estimates of the price
- We can estimate the price using only general information on the apartment
- People agree on the definition of “living room”, “bed room” etc,.

ML Model Development

- Train-test-split: standard 70-30 split
- Use pd.DataFrame.fillna() to encode all null values using one encoding (e.g., None, pd.NA, np.nan, etc.)
- Normalize numeric features using yeo-johnson power transformation
- Standarize numeric features using standard scaler
- Normalize label (price) using box-cox power transformation
- KNN Imputation

ML Models Trials

Linear regression:
RMSE: 0.93

Linear Regression with Elastic Net Regularization:
RMSE: 0.90

Hyper Tuned Decision Tree (best):
RMSE: 0.73

Good Deal Indicator

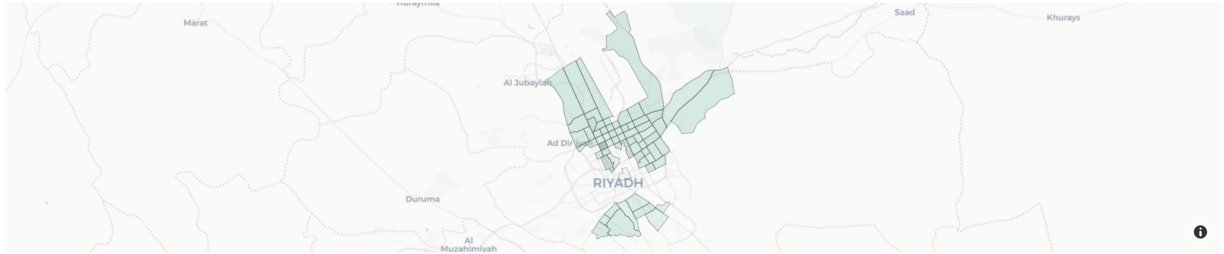
- Search Aqar's website for an ad with the specified id- scrape the data
- Clean the data
- Train an ML using a subset of our data (based on user's specification)
- Compare the model's predicted_price and real_price
- Report "deal_goodness" {Great Deal, Good Deal, Fair Deal, Poor Deal, Terrible Deal}

Goal: Help the home-buyer determine if an apartment they are interested in is fairly priced.

Required: ad_id

Explore the Sale Appartment Market in Riyadh

متوسط أسعار الشقق حسب الحي (الرياض)



Communicate: Report results in analytical dash board

Challenges

- Poor data quality
- Adequetely accurate predictions, but need to be imporoved
- Suggesting similar apartments to user’s-provided apartments
- Difficulty interpreting RMSE

Take Home Message

- Data availability and quality remain the biggest obstacles for data scientists in real estate
- It is possible to add value to the real estate industry by helping home-buyers make better-informed decisions

Future Development

- Improve data quality
- Utilize text description of properties in making price predictions
- Utilize images of properties to validate some features (e.g., furnished) and possibly more
- Utilize Google Maps-like API to expand the features in the dataset to include things like: distance_to_major_highway, distance_to_grocery_store, avg_trafic_index, etc.

THANKS!

Any questions?

Mohammed Alsalamah