

MUHAMMED UMEYR SEKBAN

ÖĞRENCİ NO: 2004040009

BÖLÜM: YAZILIM MÜHENDİSLİĞİ

MÜHENDİSLİK PROJESİ: CALİFORNİA EVLERİ

California Evleri projesi 1990’da sayımı yapılmış olup tüm California eyaletindeki evlerle ilgili bilgileri bloklar halinde incelemiştir.

Bu bilgiler şunları içerir:

1. Evlerin bulunduğu eylem ve boylam bilgileri
2. Evlerin ortalama yaşı
3. Blok başına toplam oda sayısı
4. Blok başına toplam yatak odası sayısı
5. Blok başına nüfus
6. Blok başına hane halkı
7. Blok başına ortalama gelir \$
8. Denize yakınlık

Bu bilgiler ise şu aralıklarda değişmektedir:

Boylam	Enlem	Ortalama Bina Yaşı	Toplam Oda Sayısı (Blok Başına)	Toplam Yatak Odası Sayısı (Blok Başına)
-124 / -114	32.5 / 42	1/52	2 / 39.3k	1 / 6.45k

Popülasyon (Blok Başına)	Hane Halkı (Blok Başına)	Ortalama Gelir (Blok Başına) 10k \$	Ortalama Ev Fiyatı (Blok Başına) \$
3 / 35.7k	1 / 6.08k	0.5 / 15	15k / 500k

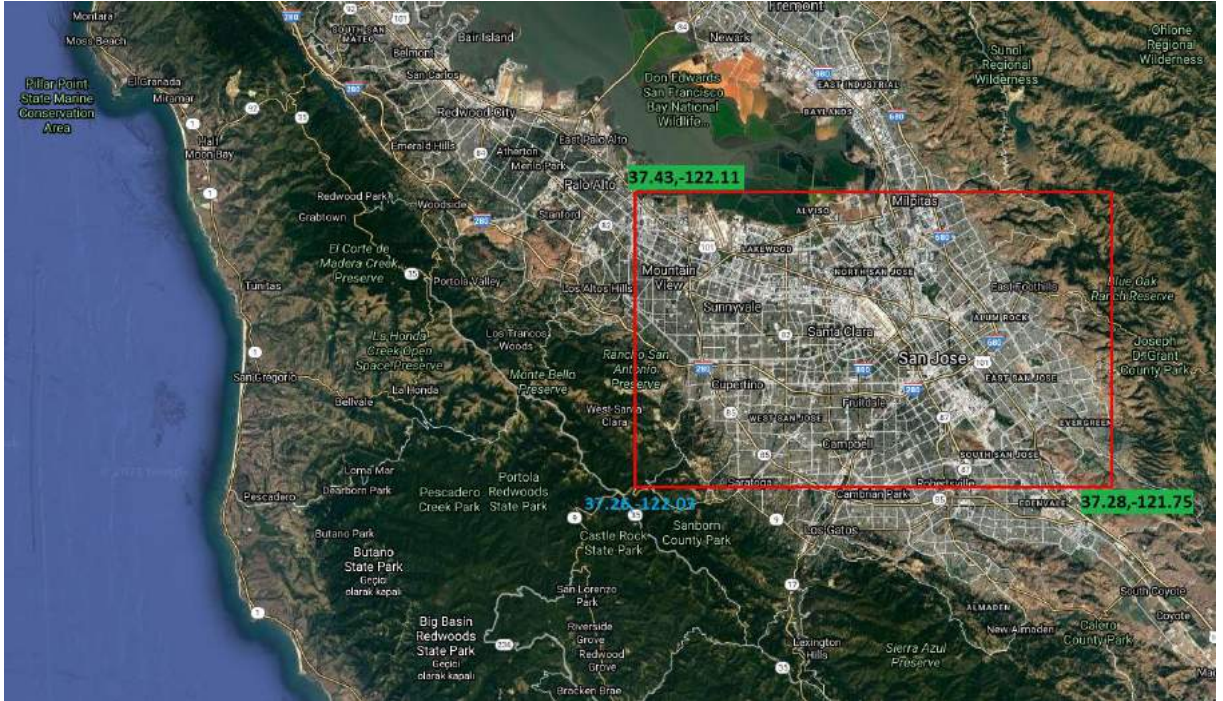
Denize Yakınlık
%44 1 saatten kısa mesafe
%32 karada
%24 diğer

Bu veriler toplamda 20 binden fazla blokun bilgisini içermektedir.

Proje kapsamında seçilen inceleme konularından birisi bu evlerin denize yakınlık/uzaklıklarına göre oda fiyatlarındaki değişimdir.

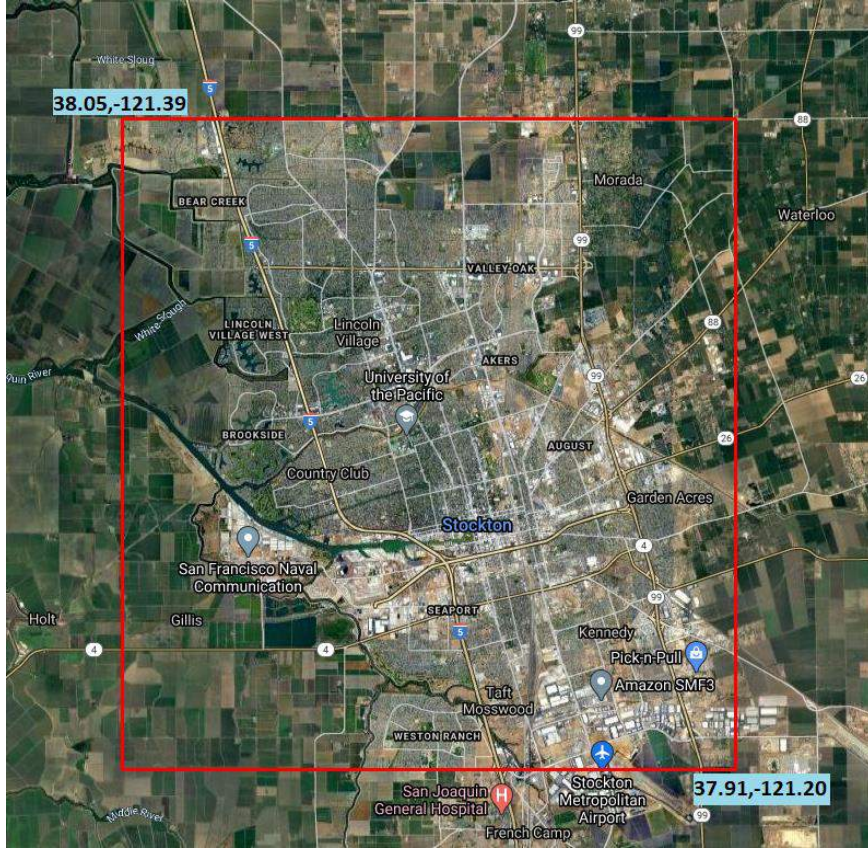
Bu değişimi incelemek için California eyaletinden belirli bölgeler seçilmiş ve bu bölgeler arasında kıyaslama yapılmıştır.

Bunlardan birincisi denize yakınlığı ortalama 35-45 mil olan **San Jose & Sunnyvale** bölgesidir.



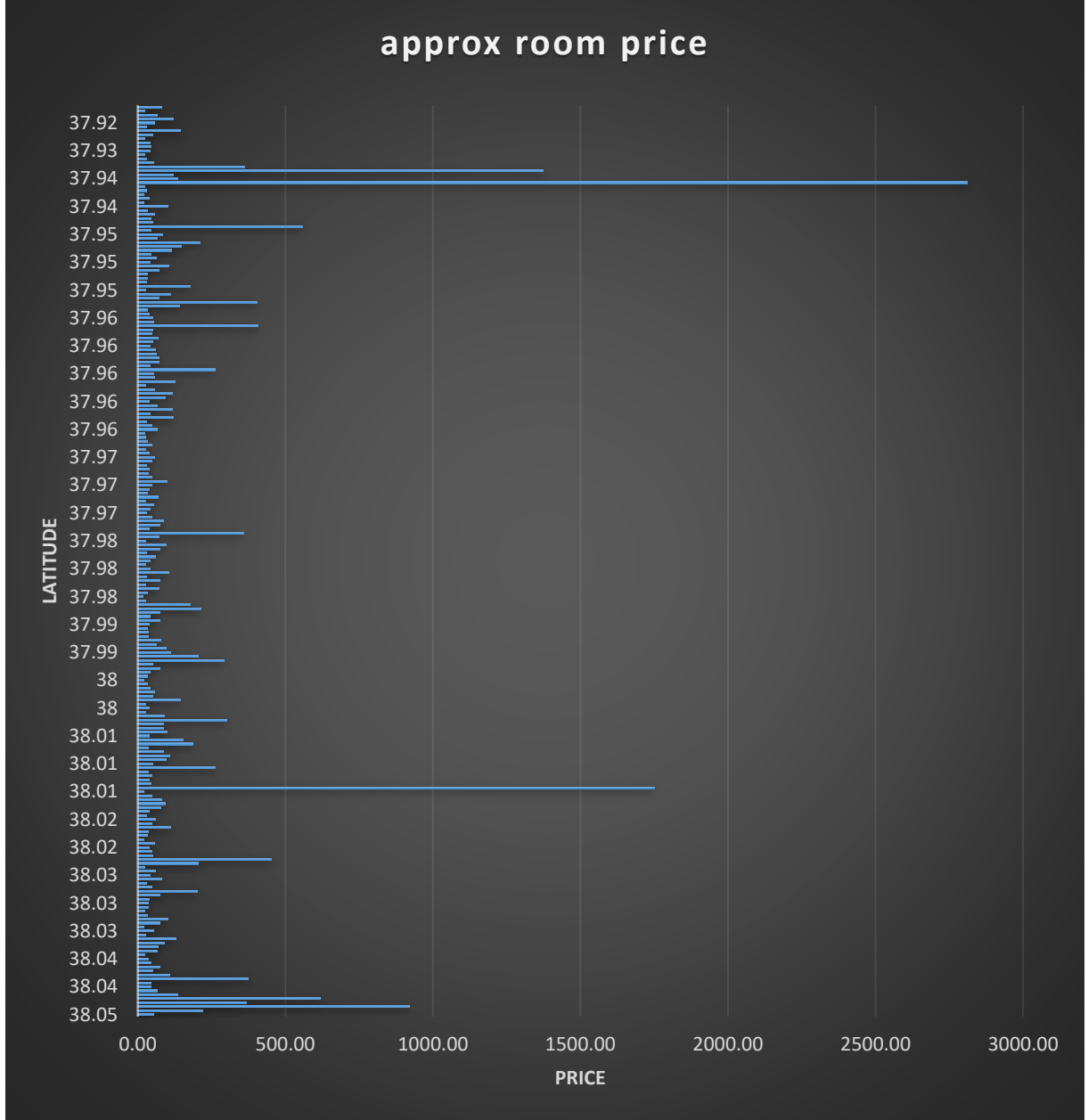
Şekil 1 Enlem ve boylam bilgileriye San Jose & Sunnyvale bölgesi

Kıyaslama yapılacak bölge ise **San Jose & Sunnyvale**’dan yaklaşık 70 mil daha doğuda yer alan **Stockton** bölgesidir.

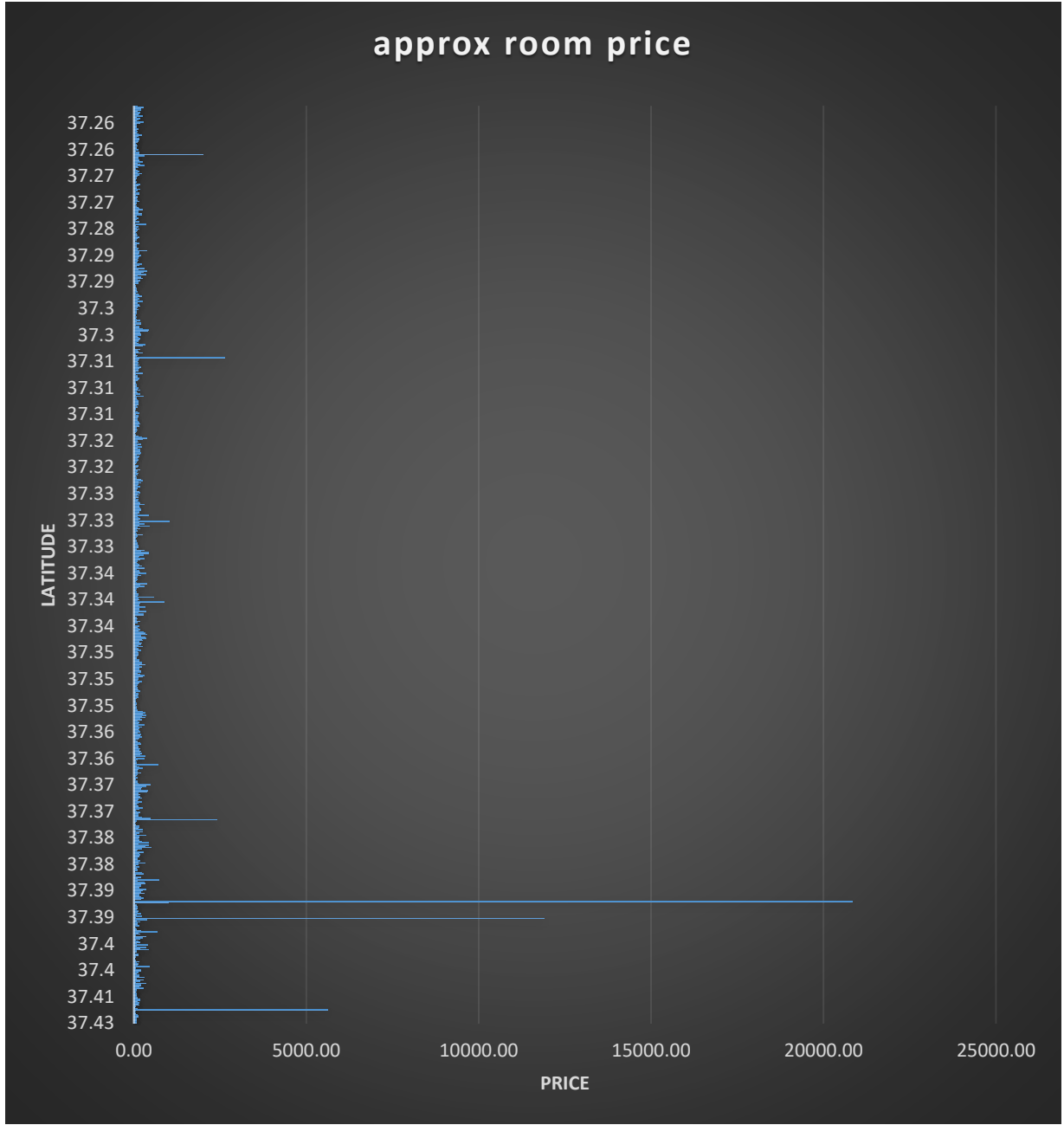


Şekil 2 Enlem ve boylam bilgileriyle Stockton bölgesi

San Jose & Sunnyvale bölgesinin kapsadığı alanda 760 blok değerlendirilirken Stockton bölgesinde 230 blok değerlendirmeye alınmıştır.



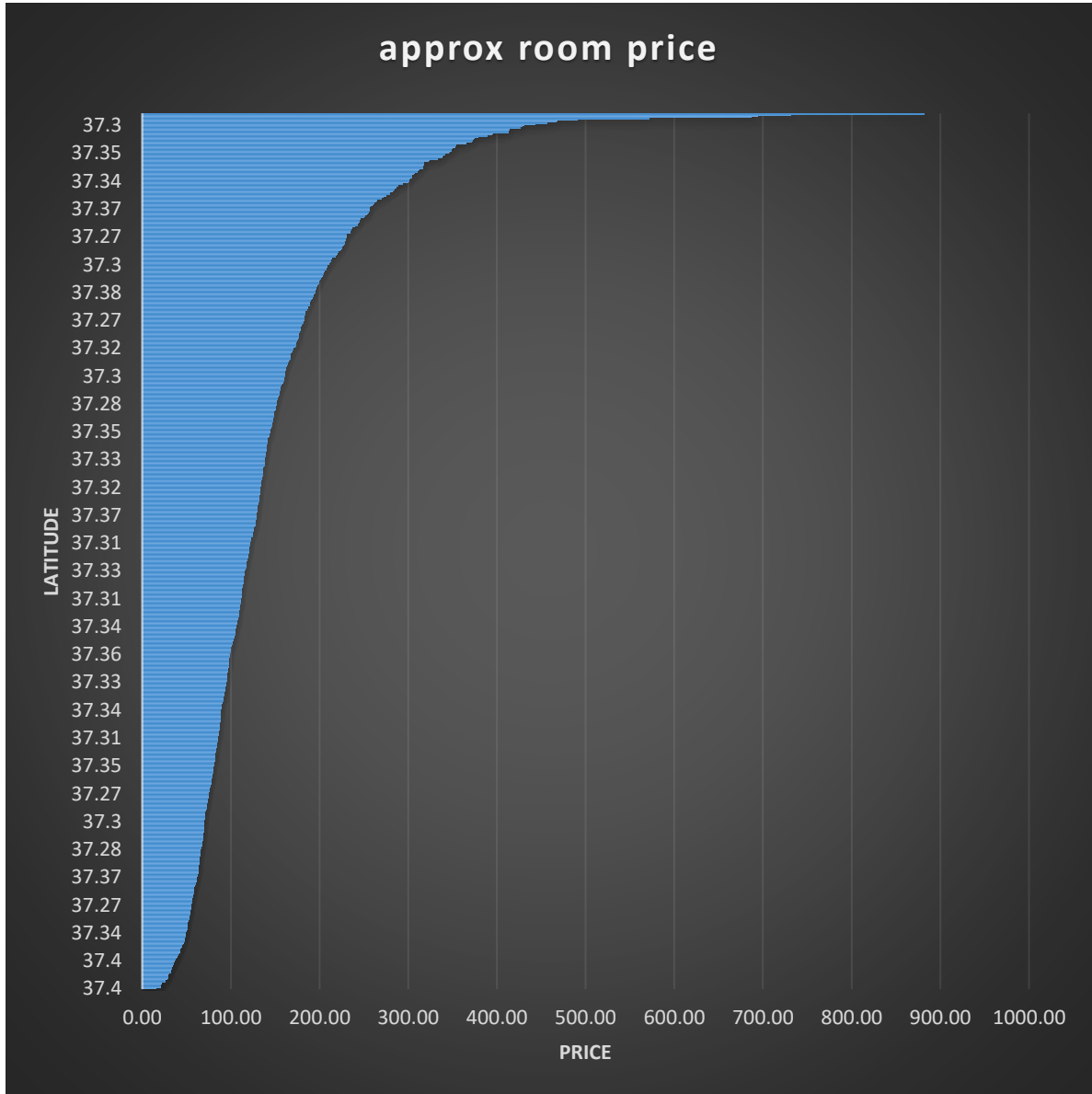
Şekil 3 Stockton bölgesinde yer alan blokların boylam bilgisine göre oda fiyatlarındaki değişim (Batıdan Doğuya)



Şekil 4 San Jose & Sunnyvale bölgesinde yer alan blokların boylam bilgisine göre oda fiyatlarındaki değişim (Batıdan Doğuya)

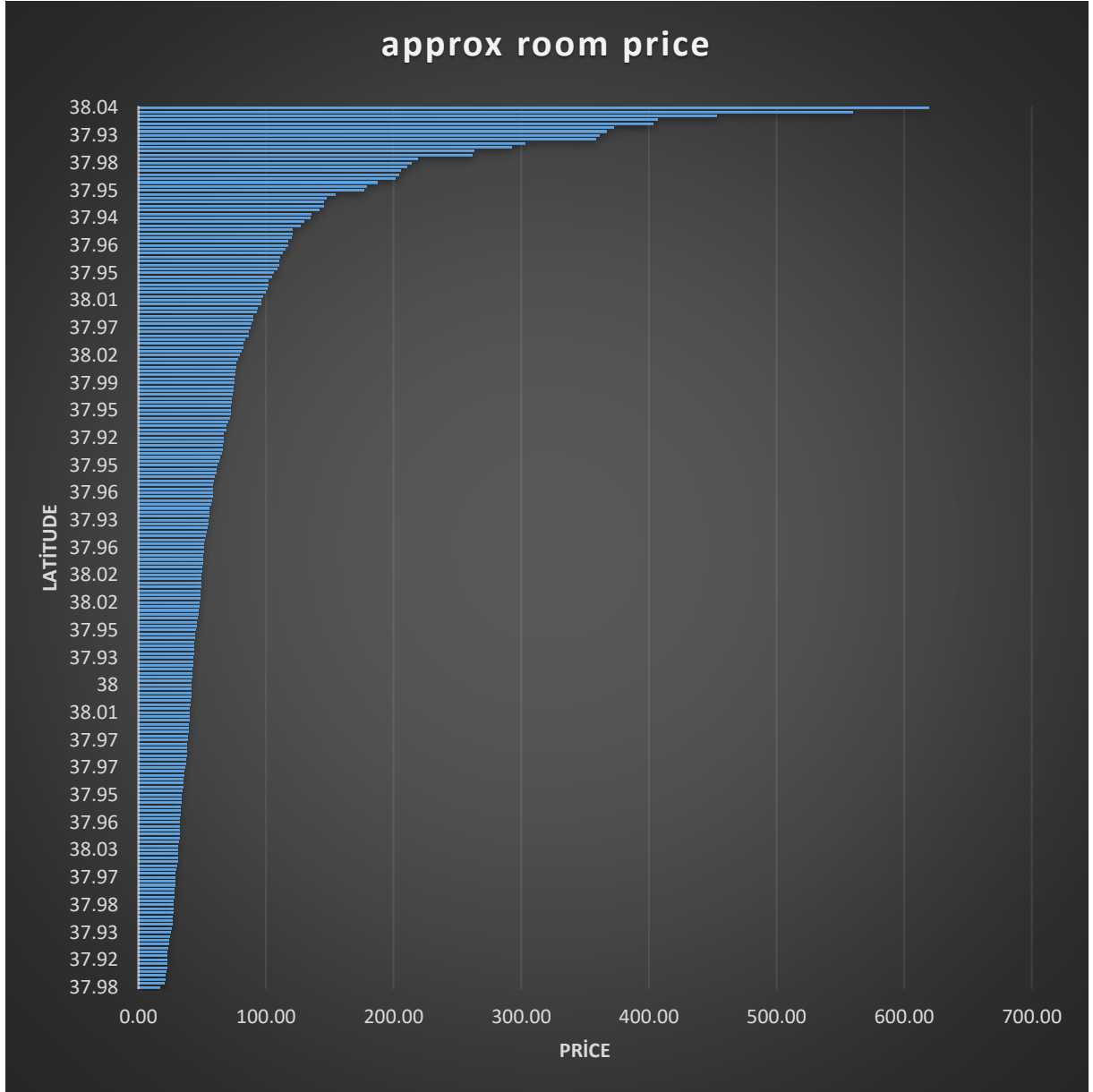
Grafiklerde görülen extrem değerler bir sonraki aşamada değerlendirmeden çıkartılarak daha sağlıklı veri incelemesi yapılacaktır.

Bu aşamada değerlendirmeye alınmayacak veriler silinerek fiyat değişimi aralığı gruplandırılmıştır.



Şekil 5 SJ & S Bölgesi, Değerler Temizlendikten Sonra

Bu fiyatlar: 0-100 arası A, 100-150 arası B, 150-250 arası C ve son olarak 250 ve daha üstü değerler D grubu olarak ayrılmıştır



Şekil 6 Stockton Bölgesi, Değerler Temizlendikten Sonra

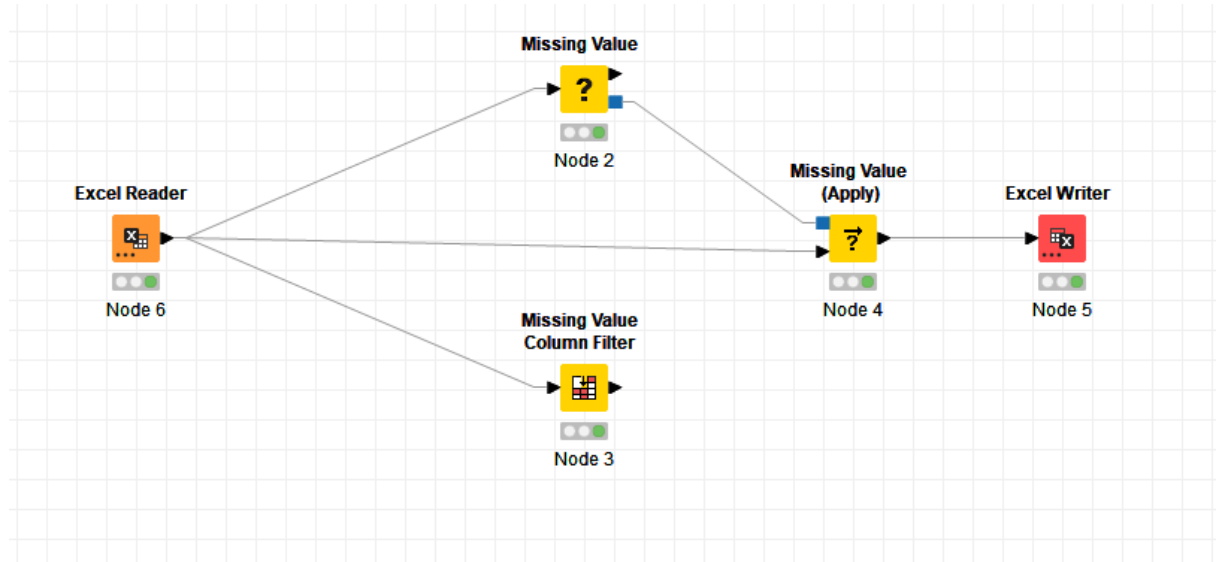
Stockton bölgesi aynı işleme tabi tutulduktan sonra gruplandırılmıştır.

0-50 arası A, 50-100 arası B, 100 ve daha yüksek değerler C grubu olarak adlandırılmıştır.

Gruplara ayrılan verilerin “yatak odası” sekmesinde eksik veriler olduğu tespit edilmiştir.

26	24	1297	297	441	282	3,1439	47500	A
27	14	6277	1742	3025	1630	4,0653	234200	A
28	9	6751		4240	1438	5,34	257400	A
29	21	4559	1163	5124	1124	3,2052	179000	A
30	17	6012	1737	3539	1625	3,8464	239400	A
31	6	2707	884	2427	804	2,6807	152400	A
488	10	1782	280	1030	237	0,4333	282200	B
489	27	2331	461	1178	447	4,6654	340700	B
490	25	2366	417	1076	398	6,9238	345900	B
491	30	1955		999	386	4,6328	287100	B
492	16	3402	1193	1479	1043	3,5861	500001	B
493	20	3401	768	1497	747	4,2188	500001	B
494	52	2009	338	841	338	5,5259	295800	B
495	31	1866	465	821	447	2,3547	275900	B
496	7	1973	328	1047	303	6,234	292200	B
497	4	2292		1050	584	4,8036	340000	B
498	36	2088	358	772	347	4,2762	310100	B
499	15	1486	310	455	296	4,3365	221000	B

Bu eksik veriler Knime ile doldurulmuştur.



Şekil 7 Yapılan İşlemin Knime Şeması

Knime’da eksik verileri doldururken ortalama değeri metodu kullanılmıştır.

Dialog - 0:2 - Missing Value

File

Default Column Settings Flow Variables Memory Policy

Number (integer) Median

Number (double) Median

String Do nothing

Options marked with an asterisk (*) will result in non-standard PMML.

OK Apply Cancel ?

Düzeltilen verileri projenin devamında işlenmek üzere yeni bir excel dosyası şeklinde çıktı alınmıştır.

13	6035	1551	2946	24	13	6035	1551	2946
21	8275	1566	3636	25	21	8275	1566	3636
24	1297	297	441	26	24	1297	297	441
14	6277	1742	3025	27	14	6277	1742	3025
9	6751	452	4240	28	9	6751		4240
21	4559	1163	5124	29	21	4559	1163	5124
17	6012	1737	3539	30	17	6012	1737	3539
6	3797	984	2437	31	6	3797	984	2437

Şekil 8 Düzeltilen verilerin karşılaştırılması -1

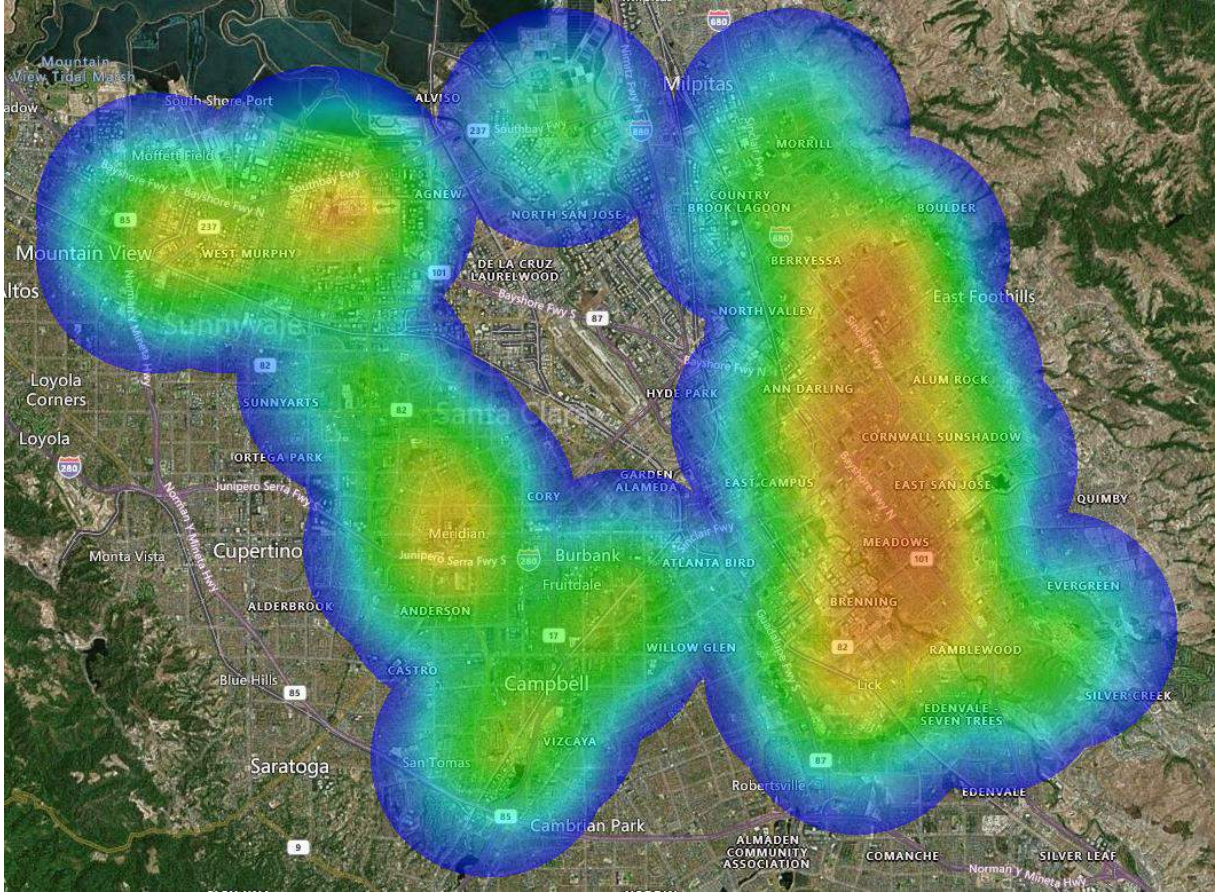
211	23	5135	911	2351	211	23	5135	911	2351
212	30	4033	794	1788	212	30	4033	794	1788
213	33	3162	452	1553	213	33	3162		1553
214	18	1525	485	1705	214	18	1525	485	1705
215	26	1934	408	2059	215	26	1934	408	2059
216	26	4310	678	1819	216	26	4310	678	1819
217	6	2850	561	2160	217	6	2850	561	2160
218	28	3085	552	1277	218	28	3085	552	1277
219	26	3756	452	1408	219	26	3756		1408
220	31	3347	589	1566	220	31	3347	589	1566
221	30	2453	573	1845	221	30	2453	573	1845

Şekil 9Düzeltilen verilerin karşılaştırılması -2

Birbirinden bağımsız ve çok farklı aralıklara dağılmış olan veriler bulunduğundan dolayı projenin bir sonraki aşamasında Knime ile Z-score normalleştirmeye gidilecektir.

Normalleştirilen veriden bu iki bölge arasında fiyat değişimine en çok eden özellik incelenecektir.

Bu özellikler incelendikten sonra proje bir emlak sitesi gibi tasarlanıp bu site üzerinden seçilen bölge, bu bölgedeki fiyat grubu ve gruptaki evlerin fiyat artışına bağlı ısı haritasına göre kullanıcıya tercih hakkı sunulacaktır.



Şekil 10 Mapcite web sitesi üzerinden elde edilen San Jose bölgesine ait fiyatların dağılımı