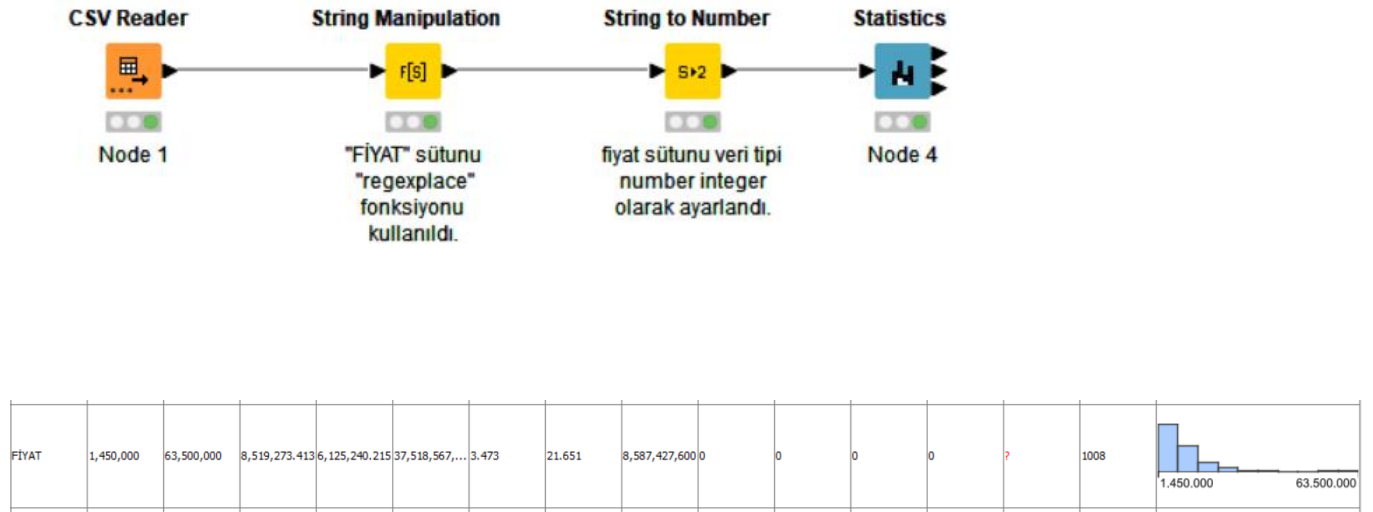


SORULAR:

1. Evlerin fiyatlarına ilişkin sütundaki değerlerin dağılımını inceleyin ve dağılımını yorumlayın. Eğer normal dağılımda değil ise, dağılımı normalleştirmek için yapılabilecek işlemleri yazınız.



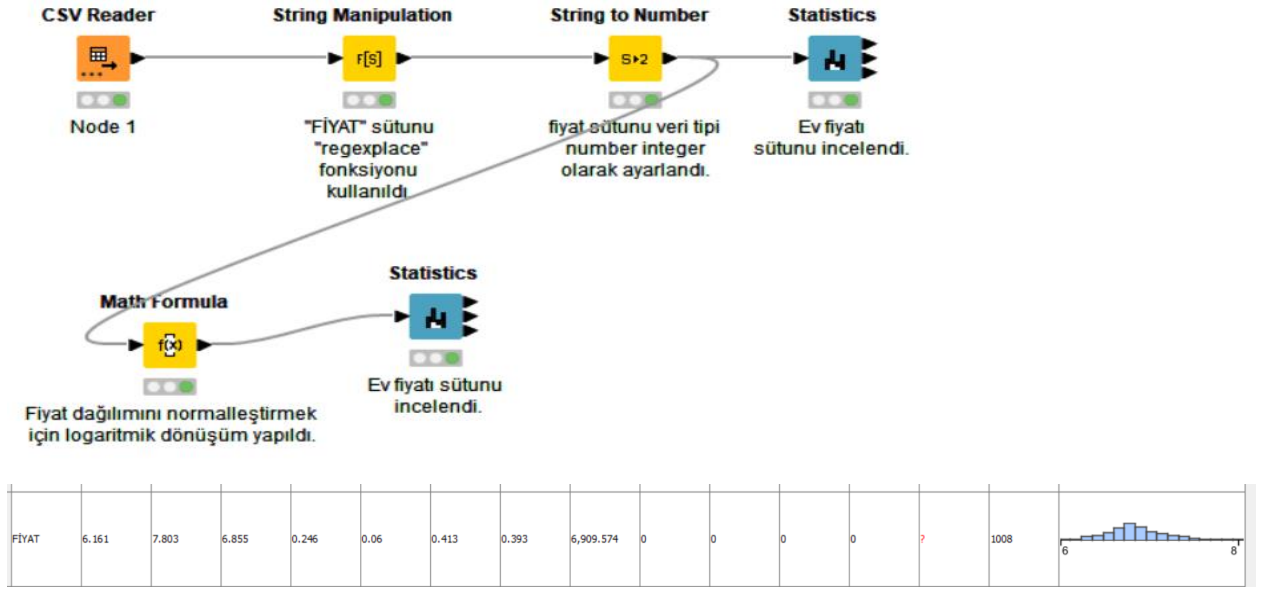
Veri setini incelemek amacıyla "CSV Reader" node'u kullanarak işleme başlanmıştır. İlk olarak fiyat sütununu inceledik ve eksik verileri kontrol ettik. Daha sonra regexReplace fonksiyonunu kullanarak fiyat sütunundaki gereksiz karakterleri kaldırdık. Bu işlemin ardından fiyat sütunu sayısal bir veri tipine dönüştürüldü.

Veri setini incelediğimizde fiyat dağılımının normal bir dağılım göstermediğini fark ettik. İstatistiksel olarak çarpıklık değeri (skewness) 3.473 olup pozitif yönde çarpık bir dağılım söz konusudur. Bu düşük fiyatların daha yoğun olduğunu ve yüksek fiyatların az sayıda ve daha geniş bir aralıkta olduğunu gösterir.

Görselleştirme: Histogram üzerinde fiyatların sağa doğru kuyruklandığı görülmektedir. Bu fiyatların normal bir dağılım sergilemediği anlamına gelir.

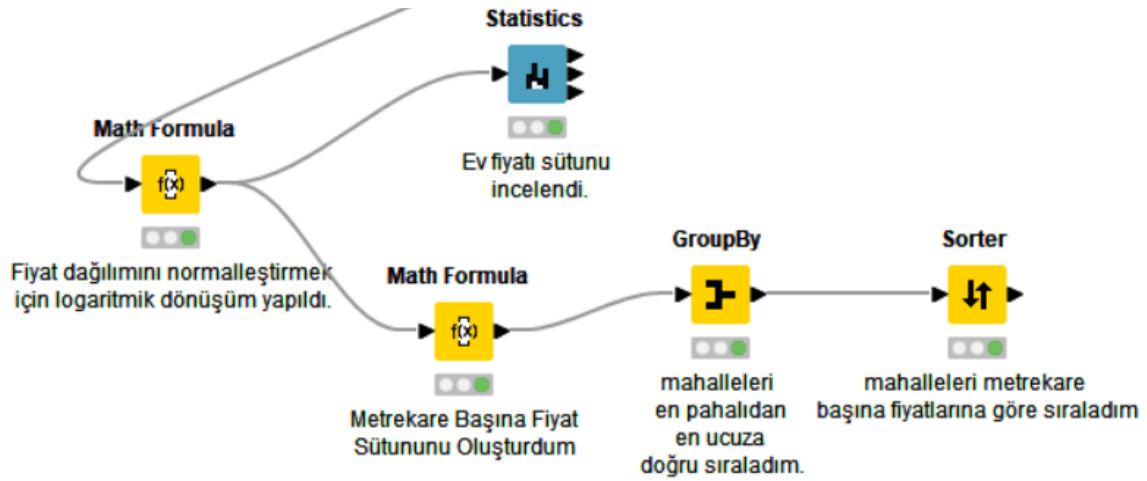
Evlerin fiyat sütununu incelediğimizde dağılımın normal olmadığını gözlemledik. Çarpıklık değeri (skewness) 3.473 olup, bu pozitif yönde bir çarpıklık olduğunu gösterir. Fiyatların büyük çoğunluğu düşük seviyelerde yoğunlaşmış olup yüksek fiyatlı evler daha nadirdir ve geniş bir aralıkta dağılmıştır.

Fiyat dağılımını normalleştirmek için yapılabilecek işlemlerden biri logaritmik dönüşümdür. Bu işlem veri setinde yer alan büyük fiyat farklılıklarını daha yakın bir seviyeye getirip dağılımı normalleştirebilir. $\log(\$FİYAT\$)$ formülü kullanarak fiyatların normal dağılıma daha yakın bir hale getirilmesine yardımcı olunabilir.



Net metrekare başına düşen fiyatın dağılımı incelendiğinde çarpıklık (skewness) değeri 0.413 olup bu dağılımın neredeyse normal bir dağılım olduğunu göstermektedir. Histogram üzerinden incelendiğinde net metrekare başına fiyatların büyük oranda dengeli bir şekilde dağıldığı görülmektedir.

2. Veri setinin içerisindeki bilgilerden yararlanarak net metrekareye düşen fiyat bakımından en pahalı olan mahalleyi bulunuz.



Veri setinin içerisindeki mahalleler metrekare başına düşen fiyat bakımından gruplanmış ve en pahalı mahalle bulunmuştur. Bu işlem için:

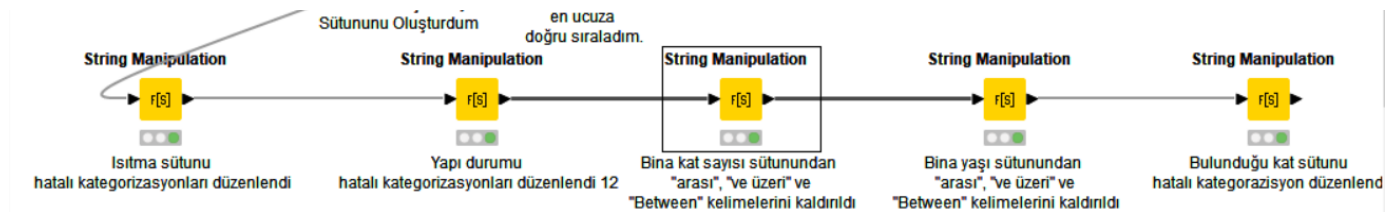
- “Group By” node’u kullanılarak mahalleler bazında net metrekare başına fiyat ortalamaları hesaplanmıştır.
- “Sorter” node’u ile mahalleler metrekare başına düşen fiyatlara göre azalan sırada sıralanmıştır.

File	Edit	Hilite	Navigation	View
Table "default" - Rows: 21				
Spec - Columns: 2		Properties		
Row ID	S MAHALLE	D	Mean(...	
Row5	Dumlupınar	0.108		
Row10	Fikirtepe	0.093		
Row16	Osmanağa	0.081		
Row7	Eğitim	0.079		
Row12	Hasanpaşa	0.078		
Row15	Merdivenköy	0.073		
Row17	Rasimpaşa	0.071		
Row20	Zühtüpaşa	0.071		
Row4	Caferağa	0.071		
Row1	Acıbadem	0.069		
Row18	Sahrayı Cedit	0.067		
Row19	Suadiye	0.066		
Row13	Kozyatağı	0.065		
Row2	Bostancı	0.065		
Row0	19 Mayıs	0.064		
Row11	Göztepe	0.063		
Row9	Feneryolu	0.062		
Row6	Erenköy	0.061		
Row8	Fenerbahçe	0.055		
Row3	Caddebostan	0.053		
Row14	Koşuyolu	0.047		

Sonuç olarak "Dumlupınar" mahallesi metrekare başına en yüksek fiyata sahip olan mahalle olarak bulunmuş ve ortalama fiyat 0.108 olarak hesaplanmıştır.

3. Aykırı veya hatalı değer tespitlerinizi varsa onlar hakkında yorumları yazınız.

Veri setinde bazı kategoriler hatalı veya tutarsız şekilde girilmişti. Bu hataları düzeltmek ve veri setinde daha tutarlı bir yapı oluşturmak için String Manipulation node'ları ile aşağıdaki işlemler gerçekleştirilmiştir:



1. Isıtma Sütunundaki Düzeltmeler:

- "Central System" ifadesi "Merkezi Sistem" olarak,
- "Combi Boiler (Natural Gas)" ifadesi "Kombi (Doğalgaz)" olarak,
- "Air Conditioner" ifadesi "Klima" olarak düzenlenmiştir.
- Ayrıca, "Soba (Doğalgaz)" ifadesi "Soba" ile değiştirilmiştir.

Bu işlemler sonucunda ısıtma sistemi kategorilerinde tutarlılık sağlanmış ve daha anlamlı hale getirilmiştir.

2. Yapı Durumu Sütunundaki Düzeltmeler:

- "Second Hand" ifadesi "İkinci El" olarak,
- "New Building" ifadesi "Sıfır" olarak değiştirilmiştir.

Böylece yapı durumu sütunundaki farklı adlandırmalar düzenlenerek kategorilerde tutarlılık sağlanmıştır.

3. Bina Kat Sayısı Sütunundaki Düzeltmeler:

- "arası", "ve üzeri", "Between" gibi ifadeler kaldırılmış ve kategoriler daha anlaşılır hale getirilmiştir.

4. Bina Yaşı Sütunundaki Düzeltmeler:

- "arası", "ve üzeri", "Between" gibi ifadeler kaldırılarak bina yaşı bilgileri daha doğru ve standart hale getirilmiştir.

5. Bulunduğu Kat Sütunundaki Düzeltmeler:

- "Raised First Floor" ifadesi "Yüksek Giriş" olarak,
- "Garden Floor" ifadesi "Bahçe Katı" olarak,
- "Zemin Kat" ifadesi "Giriş Katı" olarak değiştirilmiştir.

Bu düzenlemeler sonucunda veri setindeki hatalı ve tutarsız kategoriler düzeltilmiş, sayısal olması gereken verilerdeki stringler kaldırılmış ve veri daha anlaşılır ve analiz edilebilir hale getirilmiştir.

4. İstatistiksel olarak değerleri inceleyiniz. Değerleri yorumlayınız.

Veri setindeki sayısal sütunların dağılımları incelendiğinde bazı önemli bulgular öne çıkmaktadır. Net Metrekare sütununda ortalama 120.398 m² olarak hesaplanmış olup dağılım sağa çarpıktır (çarpıklık değeri 1.839). Bu veri setinde birkaç büyük metrekareli dairenin genel ortalamayı etkilediğini göstermektedir.

Brüt Metrekare sütunu ise oldukça geniş bir aralıkta daireler içermekte ortalama 152.492 m² olup dağılımın çok yüksek çarpıklık gösterdiği (4.35) görülmektedir. Bu veri setinde aşırı büyük değerlerin bulunduğunu işaret etmektedir.

Oda Sayısı sütunu daha dengeli bir dağılım göstermekte olup ortalama 3 odalı dairelerin yaygın olduğunu Salon Sayısı sütununda ise yüksek çarpıklık (3.193) bulunmakta bu da genellikle 1 salonlu dairelerin ağırlıkta olduğunu göstermektedir.

Son olarak Aidat sütununda ortalama 93.945 TL değerinde aidat bedelleri tespit edilmiş ancak yüksek standart sapma ve çarpıklık değeri (2.29) bazı dairelerde oldukça yüksek aidat bedellerinin mevcut olduğunu ortaya koymaktadır.

Bu bulgular veri setinde özellikle büyük dairelerin ve yüksek aidat bedellerinin genel dağılımı etkilediğini göstermektedir.