پروژه درس مبانی داده کاوی

فاز اول

علی جعفری (9622762431)

سوال اول:

با توجه به کد و مشاهده دیتاست دیوار، جدول خواسته شده به صورت زیر پر شده است. بازه مقادیر، به علت طولانی بودن، در پایین جدول و تنها برای بعضی از ویژگی ها آورده شده است. زیرا بعضی از ویژگی ها (مانند id)، خاصیت یکتا بودن داشته و تعداد مقادیر مختلف آنها بسیار زیاد میشود. مقادیر پرت نیز تنها برای ویژگی های از نوع Numeric محاسبه شده است، زیرا همانطور که مشخص است، مقادیری مانند مینیمم، ماکسیمم و Median که برای رسم Box Plot به آنها نیاز است، در این دیتاست تنها برای مقادیر عددی تعریف شده و با معنی هستند. مقادیری هم که در جدول با خط تیره پر شده اند، به این نشانه هستند که آن مشخصه خاص، برای آن ویژگی خاص، بی معنی است. همچنین، از آوردن بازه مقادیر برای ویژگی های عددی پرهیز شده است. زیرا همان طور که مشخص است، مقادیر مینیمم و ماکسیمم این ویژگی ها اندازه گیری شده و بازه مقادیر ویژگی های عددی را میتوان با استفاده از مینیمم و ماکسیمم این ویژگی ها، بازنمایی کرد.

|  |  |  |  |  |  |  |  |
| --- | --- | --- | --- | --- | --- | --- | --- |
| ردیف | نام ویژگی | نوع | Min | Max | Mean | Mode | Median |
| 1 | archive\_by\_user | Symmetric Binary | - | - | - | True | - |
| 2 | brand | Nominal | - | - | - | پراید صندوق دار | - |
| 3 | cat1 | Nominal | - | - | - | for-the-home | - |
| 4 | cat2 | Nominal | - | - | - | furniture-and-home-decore | - |
| 5 | cat3 | Nominal | - | - | - | light | - |
| 6 | city | Nominal | - | - | - | Tehran | - |
| 7 | created\_at | Nominal | - | - | - | Saturday 12PM | - |
| 8 | desc | Nominal | - | - | - | کاملا سالم | - |
| 9 | id | Nominal | - | - | - | - | - |
| 10 | image\_count | Numeric | 0 | 16 | 1.78 | 0, 1 | 2 |
| 11 | mileage | Numeric | 0 | 500,000 | 100862.29 | 200000 | 75,000 |
| 12 | platform | Symmetric Binary | - | - | - | mobile | - |
| 13 | price | Numeric | 1 | 400,000,000 | 4,096,271.89 | 100,000 | 220,000 |
| 14 | title | Nominal | - | - | - | بوفه | - |
| 15 | type | Nominal | - | - | - | women | - |
| 16 | year | Numeric | 1366 | 1395 | 1387.64 | 1393 | 1389 |

**بازه مقادیر:**

1- archive\_by\_user: False, True

2-