In [1]:
```
# import packages
import pandas as pd
from pandas_profiling import ProfileReport
```

```
C:\Users\omgha\AppData\Local\Temp\ipykernel_22748\3122818937.py:3: Deprecatio
nWarning: `import pandas_profiling` is going to be deprecated by April 1st. P
lease use `import ydata_profiling` instead.
  from pandas_profiling import ProfileReport
```

In [2]:
```
# Read the file we have
df1 = pd.read_csv('quikr_car.csv')
df1
```

Out[2]:

	Unnamed: 0	name	company	year	Price	kms_driven	fuel_type
0	0	Hyundai Santro Xing	Hyundai	2007	80000	45000	Petrol
1	1	Mahindra Jeep CL550	Mahindra	2006	425000	40	Diesel
2	2	Hyundai Grand i10	Hyundai	2014	325000	28000	Petrol
3	3	Ford EcoSport Titanium	Ford	2014	575000	36000	Diesel
4	4	Ford Figo	Ford	2012	175000	41000	Diesel
...	...	...	...	...	...	...	...
811	811	Maruti Suzuki Ritz	Maruti	2011	270000	50000	Petrol
812	812	Tata Indica V2	Tata	2009	110000	30000	Diesel
813	813	Toyota Corolla Altis	Toyota	2009	300000	132000	Petrol
814	814	Tata Zest XM	Tata	2018	260000	27000	Diesel
815	815	Mahindra Quanto C8	Mahindra	2013	390000	40000	Diesel

816 rows × 7 columns

In [3]:
```
# genrate report
profile = ProfileReport(df1)
profile.to_file(output_file= "Car_report.html")
profile
```

Summarize dataset: 100% 32/32 [00:02<00:00, 14.24it/s, Completed]

Generate report structure: 100% 1/1 [00:01<00:00, 1.85s/it]

Render HTML: 100% 1/1 [00:01<00:00, 1.55s/it]

Export report to file: 100% 1/1 [00:00<00:00, 24.48it/s]

Overview

Dataset statistics

Number of variables	7
Number of observations	816
Missing cells	0
Missing cells (%)	0.0%
Duplicate rows	0
Duplicate rows (%)	0.0%
Total size in memory	44.8 KiB
Average record size in memory	56.2 B

Variable types

Numeric	4
Categorical	3

Alerts

`name` has a high cardinality: 254 distinct values	**High cardinality**
`year` is highly overall correlated with `Price`	**High correlation**
`Price` is highly overall correlated with `year` and 1 other fields (year, company)	**High correlation**

Out[3]: