Machine Learning Project

Բնակարանների Գների Կանխատեսում

Գծային Ռեգրեսիայի (Linear Regression) մոդելի կառուցում և տվյալների վերլուծություն Python-ի միջոցով

1. Տվյալների Նախապատրաստում

Pandas գրադարանով տվյալների աղյուսակի ստեղծում և նախնական ուսումնասիրություն

Տվյալների Կառուցվածքը (DataFrame)

☑ Մուտքային Հատկանիշներ (X)

- **area:** Բնակարանի մակերեսը ($մ^2$)
- **bedrooms:** Ննջասենյակների քանակը
- **age:** Շենքի տարիքը (տարիներով)
- **location_score:** Տեղադրության միավորը (1-10)

◎ Թիրախային Փոփոխական (y)

- **price:** Բնակարանի գինը (հազար $)
- Հանդիսանում է կանխատեսվող հիմնական ցուցանիշը
- Կախվածության մեջ է գտնվում բոլոր 4 մուտքային պարամետրերից

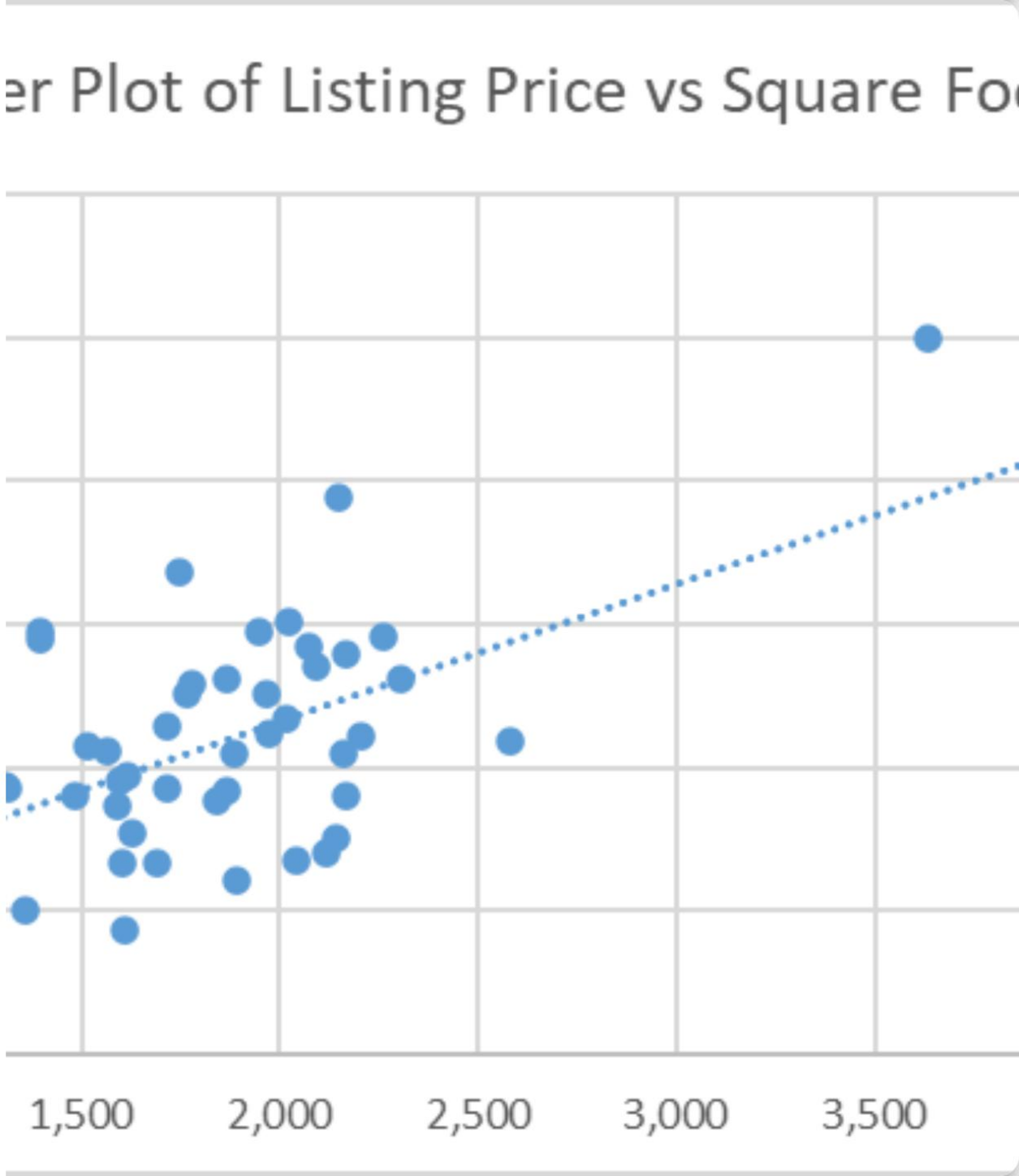
Տվյալների Վիզուալիզացիա

Area vs Price Կախվածությունը

Matplotlib գրադարանի **plt.scatter()** ֆունկցիայով կառուցվել է կետային գրաֆիկ:

Գրաֆիկը ցույց է տալիս ուղիղ համեմատական կապ. մակերեսի ավելացմանը զուգընթաց բնակարանի գինը օրինաչափորեն աճում է:

Սա հաստատում է գծային մոդելի կիրառման արդյունավետությունը:



Կորելյացիոն Վերլուծություն

df.corr()

Հաշվում է փոփոխականների միջև կորելյացիոն գործակիցները (-1-ից 1 միջակայքում):

Seaborn Heatmap

sns.heatmap()-ը պատկերում է ջերմային քարտեզ, որը տեսանելի է դարձնում կապերի ուժգնությունը:

Հիմնական Եզրահանգում

Մակերեսը (Area) և տեղադրությունը (Location) ունեն ամենաբարձր դրական կորելյացիան գնի հետ:

2. Մոդելի Կառուցում և Ուսուցում

Scikit-Learn գրադարանի կիրառումը մեքենայական ուսուցման գործընթացում

Train / Test Split

80/20

Train / Test Հարաբերակցություն

train_test_split()

X_train, y_train (80%): Օգտագործվում է մոդելին ուսուցանելու և կանոնները սովորեցնելու համար:
X_test, y_test (20%): Պահվում է առանձին՝ մոդելի իրական ճշտությունը ստուգելու համար:
random_state=42: Ապահովում է պատահականության կրկնելիությունը:

Գծային Ռեգրեսիայի Մոդելը

⚙️ model.fit()

Ուսուցման փուլ: Մոդելը վերլուծում է ուսուցողական տվյալները (X_train, y_train) և գտնում է լավագույն մաթեմատիկական բանաձևի կոեֆիցիենտները:

🪄 model.predict()

Կանխատեսման փուլ: Մոդելը ստանում է թեստային բնակարանների տվյալները (X_test) և հաշվում է դրանց ենթադրյալ գները (y_pred):

Արդյունքների Համեմատություն

Բնակարան	Մակերես ($մ^2$)	Իրական Գին (y_test)	Կանխատեսված Գին (y_pred)
Բնակարան 1 (Index 1)	60 $մ^2$	$125,000	$124,150
Բնակարան 2 (Index 7)	140 $մ^2$	$300,000	$298,800

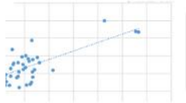
Գծային ռեգրեսիայի մոդելը ցուցադրում է բարձր ճշտություն՝ minimale շեղումներով:

Հարցեր և Պատասխաններ

Շնորհակալություն ուշադրության համար

Machine Learning Project | Python Data Science

Image Sources



https://media.cheggcdn.com/media/333/3330dcd6-8411-4188-91da-5ffc673ef218/phpvD2YyN

Source: www.chegg.com