

Data Analysis

Analysis was carried out to investigate the effect of information revelation and genre of the popularity of books. The linear regression model consisting of average, standard deviation, range and the slope of a linear regression across the course of the narrative did not explain much of the variation in the popularity (adj r-square = 0.04). The addition of the genre variable to the regression did not significantly improve this. The standard deviation range as and slope were significant predictors in the model with the higher slopes resulting in lower popularity levels.

The Lasso regression further supported the findings from the linear regression as the coefficients were all zero showing that they were less significant predictors in the model. A negative R squared value obtained in the lasso regression also point to the poor fit of the model.

Linear Regression Output without the Genre

OLS Regression Results

Dep. Variable:downloads_newR-squared:0.004

Model:OLSAdj. R-squared:0.004

Method:Least SquaresF-statistic:18.85

Date:Fri, 05 Jul 2024Prob (F-statistic):1.75e-15

Time:15:24:59Log-Likelihood:-29353.

No. Observations:18988AIC:5.872e+04

Df Residuals:18983BIC:5.876e+04

Df Model:4

Covariance Type:nonrobust

	coef	std err	t	P> t	[0.025	0.975]
const	3.2274	0.132	24.400	0.000	2.968	3.487
avg_kld	-0.0540	0.609	-0.089	0.929	-1.248	1.140
std_dev_kld	8.3763	1.744	4.802	0.000	4.957	11.795
range_kld	-1.3876	0.245	-5.675	0.000	-1.867	-0.908
slope_kld	-75.9182	16.035	-4.735	0.000	-107.348	-44.488
genre_Historical	0	0	nan	nan	0	0
genre_Non-Fiction	0	0	nan	nan	0	0
genre_Other	0	0	nan	nan	0	0
genre_Others	0	0	nan	nan	0	0
genre_Science	0	0	nan	nan	0	0
genre_other	0	0	nan	nan	0	0

Omnibus:2532.904Durbin-Watson:1.633

Prob(Omnibus):0.000Jarque-Bera (JB):5133.008

Skew:0.831Prob(JB):0.00

Kurtosis:4.931Cond. No.inf

Variables: The variables include the average, standard deviation, range and slope.

Lasso Regression Coefficient Output

	Variable	Coefficient
0	avg_kld	-0.0
1	std_dev_kld	-0.0
2	range_kld	-0.0
3	slope_kld	-0.0
4	genre_Historical	0.0
5	genre_Non-Fiction	0.0
6	genre_Other	0.0
7	genre_Others	0.0
8	genre_Science	0.0
9	genre_other	0.0

Variables

1. Genre: The genre was obtained using the NLP tf-idf procedure. The subjects were processed to make it easy to identify common words, which were then ranked. Keywords that are common to genres were developed from online sources as shown below. , and thus were the used in identifying the groups to which the various books are likely to belong to.

```
+ Code + Text
[100] 'Fiction': [
    'imagination', 'narrative', 'characters', 'plot', 'story', 'literature', 'prose', 'fantasy', 'romance',
    'adventure', 'thriller', 'suspense', 'mystery', 'fiction', 'novel', 'drama', 'epic', 'saga', 'allegory',
    'myth', 'fable', 'novella', 'storytelling', 'character-driven', 'plot-driven', 'short story', 'fictional',
    'sea stories', 'domestic', 'political', 'legal thriller', 'science fiction', 'utopian',
    'dystopian', 'war stories', 'love stories', 'satire', 'gothic', 'pastoral', 'tragedies', 'didactic',
    'bildungsromans', 'humorous', 'ghost', 'fairy tales', 'folklore', 'paradoxical', 'paranormal'
],
'Non-Fiction': [
    'factual', 'real', 'informative', 'documentary', 'true story', 'educational', 'biography', 'memoir', 'essay',
    'reference', 'analysis', 'research', 'non-fiction', 'guide', 'manual', 'how-to', 'textbook', 'instructional',
    'scholarly', 'report', 'non-fictional', 'journal', 'chronicle', 'historical account', 'autobiography',
    'essays', 'poetry', 'epics', 'handbooks', 'manuals', 'sources', 'speeches', 'addresses', 'political science',
    'geography', 'methodology'
],
'Historical': [
    'history', 'period', 'past', 'era', 'heritage', 'historical events', 'authenticity', 'ancient', 'medieval',
    'reconstruction', 'civilization', 'chronicle', 'historical', 'antiquity', 'archaeological', 'cultural',
    'traditional', 'historic', 'bygone', 'retrospective', 'historical narrative', 'historiography', 'timeline',
    'historical fiction', 'revolution', 'civil war', 'colonial', 'new plymouth', 'foreign relations',
    'population statistics', 'political history', 'constitutional history', 'inaugural addresses', 'census'
],
'Science': [
    'research', 'discovery', 'experiment', 'theory', 'data', 'analysis', 'biology', 'chemistry', 'physics',
    'innovation', 'technology', 'scientific', 'exploration', 'technical', 'medical', 'environmental', 'engineering',
    'biological', 'chemical', 'physical', 'scientific discovery', 'technological advancement', 'experimental',
    'laboratory', 'scientific method', 'internet', 'computer networks', 'information retrieval', 'data processing',
    'electronic publishing', 'text processing', 'computer security', 'radio broadcasting', 'mathematical constants',
    'number theory'
],
'Other': [
    'self-help', 'motivation', 'improvement', 'success', 'advice', 'wellness', 'personal growth', 'mental health',
    'habits', 'coaching', 'guidance', 'humor', 'poetry', 'drama', 'art', 'philosophy', 'self-development',
    'self-improvement', 'motivational', 'inspirational', 'self-care', 'self-love', 'mindfulness', 'meditation',
    'spirituality', 'leadership', 'creativity', 'satire', 'parody', 'comedic', 'artistic', 'philosophical',
    'reflective', 'children's literature', 'young adult', 'juvenile', 'fairy tales', 'sacred books', 'bible',
    'christian', 'mythology', 'folklore', 'epic poetry', 'greek drama', 'drama adaptations', 'magical realism',
    'urban fantasy', 'paranormal romance', 'imaginary places', 'mythological characters'
]
```