Site Acquisition Analysis: Companion's Breweries

Executive Summary

Brewery Companion is seeking to improve its competitiveness through the acquisition of either site A or site B. The objective is to make a data-driven decision on the basis of a predictive equation model.

This report presents an analysis of historical data from Companion's 47 pubs, including variables such as revenue (REV), management quality (MAN), staff quality (STA), interior type (TYP), annual per capita income (INCOME) and competitor density (COM). In the initial data analysis, we examined the scatterplots of each variable in relation to the independent variable REV and assessed the correlations. It is evident that there is a clear linear relationship between MAN and REV, with no evidence of multicollinearity.

Backward elimination regression analysis was used to construct the predictive equation model. The results identified MAN, TYPE and INCOME as statistically significant variables. Analysis of the residual scatter plots indicated that the variable INCOME was non-linear and therefore a predictive model with a quadratic term for income was used. The model is robust and gives an adjusted R-squared value of 98.7%.

Based on our assumption of different quality of management between the sites, site evaluation suggests that Site A offers a more stable income outlook (£4.20m to £12.81m) than Site B (-£2.01m to £7.35m). To improve forecast accuracy, we recommend collecting additional data on sales types, demographics and customer profiles. Management quality remains key, explaining around 75% of the variance in revenue. Nevertheless, given the stability of the revenue forecast and the potential for model refinement, we recommend prioritizing the acquisition of Site A.

In summary, Companion should focus on acquiring Site A, emphasizing the importance of exceptional management and future data collection to improve the analysis.

Introduction

The UK brewing industry landscape has undergone significant shifts, with international competitors vying for market share and regulatory bodies scrutinizing the trade. To stay competitive, Companion's is embarking on a strategic analysis of our existing pubs to discern the key factors influencing revenue. We will be specifically evaluating external growth through the metric of revenue potential, not profitability.

In the "Findings" section of the report, we examine data from our current establishments, including evaluations of management and staff quality, a categorization of establishment type, local per capita annual income, and an index reflecting competitor density. Our analysis is based on known referentials and historical data, with the understanding that force majeure events, or major crises, have not been factored in our acquisition assessment.

This report evaluates two prospective sites (Sites A and B) offered to Companion's, providing recommendations on their potential acquisition based on the established prediction equation from the five categories listed above. The two sites differ with respect to their location (rural vs. urban), interior (rustic vs. modern), clientele (local vs. diverse), and staff/management suitability.

Findings

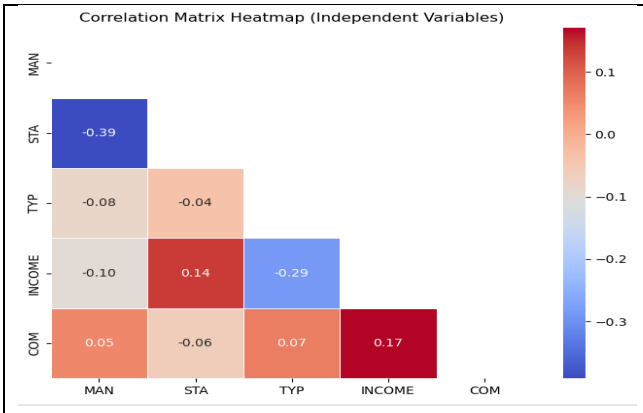
Initial Data Analysis

In order to make a data driven decision which site to acquire we analyzed the following available historical data for revenue (REV), quality of management (MAN), quality of staff (STA), interior type (TYP), per capita annual income in the region (INCOME) and number of competitors per squared kilometer (COM) for 47 pubs owned by Companion's Breweries. Summary statistics of the dataset are available in Table 1 below.

	REV (£ millions)	**MAN**	**STA**	**TYP**	**INCOME (**£000s)	**COM**
count	47	47	47	47	47	47
mean	8.69	5.30	5.74	0.45	10.68	0.64
std	3.95	2.28	1.86	0.50	1.89	0.21
min	0.49	1.00	1.00	0.00	7.00	0.10
25%	6.02	4.00	5.00	0.00	9.00	0.51
50%	9.38	5.00	6.00	0.00	10.20	0.67
75%	11.06	7.00	7.00	1.00	12.15	0.79
max	17.65	9.00	9.00	1.00	16.00	0.98

Table 1: Summary statistics of the analyzed dataset.

In order to determine potential interdependencies among the explanatory variables, we calculated a correlation matrix.



Correlation matrix.

The correlation matrix shows a low to moderate negative correlation between MAN and STA, indicating that in our sample higher evaluations of management tend to be associated with low evaluation of staff (and vice versa). Also, we observe low negative correlation between INCOME and TYP, indicating that in our sample pubs with traditional interiors (TYP=0) tend to be in areas with higher annual income per capita.

Correlations among other variables are very low and, given the magnitude of all correlations, we can **exclude multicollinearity** among explanatory variables in our sample.

Next, we plotted all independent variables against the value of REV in order to explore their relationship with revenue (illustrated in Figure 1, Scatter Plots).

Qualitatively, we can observe a strong positive linear dependence between MAN and REV, an indication of positive linear dependence between REV and TYP, and an indication of negative linear dependence between REV and STA and between REV and INCOME. We do not observe particular outliers that may affect a linear regression against the explanatory variables.

In the rest of our analysis, in the absence of further information, we assume that each explanatory variable was measured independently of revenue. This is particularly relevant for variables that may be affected by subjective judgement, like quality of management and of staff. We assume that quality of management is measured in the same way for all pubs, with no consideration of revenue of the pub, through tools like 360-degree feedback from co-workers, type and amount of education and trainings, job experience, etc. We also assume that quality of staff is measured using a similar approach, identical across different pubs, using the same scale, so that measurements of STA in different pubs are comparable.

Regression Analysis (Backward Elimination)

We conducted a stepwise backward regression, initially including all five categories in our model, subsequently removing the least significant predictors until we identified the most influential ones for our model.

STEP 1: 5 VARIABLES

This regression model (Figure 2, 5 Variables) has a high Adjusted R-squared value (94.5%), indicating that most of the variance in revenue is explained by the independent variables. STA and COM, however, are statistically insignificant, as demonstrated by their high p-value. This outcome for STA was expected due to the noteworthy collinear relationship between STA and MAN.

STEP 2: 4 VARIABLES

In the subsequent model (Figure 3, 4 Variables), COM (least statistically significant variable) was eliminated, resulting in an increased Adjusted R Square (94.6%). STA remains statistically insignificant with a high p-value ($p > 0.05$).

STEP 3: 3 VARIABLES

In our third model (Figure 4, 3 Variables), the Adjusted R Square remains high (94.7%), indicating a good fit. The remaining variables, MAN, TYP and INCOME are all statistically significant predictors of revenue with a p-value less than 5%.

These coefficients represent the estimated impact of each variable on revenue. For example, a one-unit increase in MAN is associated with an increase of approximately 1.5989 in revenue, holding other variables constant.

The residual plot for INCOME (Figure 5, Residual Plot for INCOME), however, exhibits a non-linear relationship with the dependent variable REV, calling for an additional assessment of this independent variable.

STEP 4: Quadratic Variable

To overcome the non-linearity in INCOME, we concluded that running an adjusted model with another variable, i.e. INCOME_SQ (Figure 6, Quadratic Term), will improve our results. By so doing, we observe an increase in the Adjusted R Square (98.7%), which means that INCOME_SQ has increased the explanatory power of the model. Furthermore, coefficients are statistically significant and residual plots are randomly scattered around the x-axis.

Recommendation

In our analysis of Companion's Breweries' decision regarding the acquisition of two potential sites, Site A and Site B, we employed the multiple regression model with the quadratic term included to predict the revenues for each site based on their specific characteristics. This predictive model has been derived after conducting the backward elimination process. The predictive equation comes from the results of our multiple regression model and is summarized below:

Prediction Equation:

Predicted Revenue (in millions of £) =

$26.0577 + (1.5126 * MAN) + (1.1021 * TYP) + (-4.3543 * INCOME) + (0.1755 * INCOME\_SQ^2) + e$

Site Characteristics:

MAN_A = 4 (We assume that site A will have a better management quality score because, although the owners, who we assume to be the current management, wish to retire, the staff currently employed there are trustworthy and loyal and, as they are most likely to stay, they know

the business and the customers, we assume that there is a possible candidate for management. However, we have not given a high score as we would still need to provide the candidate with enhanced tools and insights).

MAN_B = 1 (For Site B it will be difficult to find the right manager, so we assume a lower management score).

TYP_A = 0 (We assume that site A will have a Conventional type of pub because we have the info that it is pleasant but in need of minor repairs and re-thatching and is in a small village with a lot of agricultural workers).

TYP_B = 1 (We assume that site B will have a Modern or unusual theme/ambience pub as the pub is a converted disused inner-city church).

INCOME_A = 8 (Local per capita annual income for Site A in millions of £).

INCOME_B = 14.8 (Local per capita annual income for Site B in millions £).

INCOME_SQ_A = INCOME_A ^ 2

INCOME_SQ_B = INCOME_B ^ 2

Results:

1) Predicted Revenue for Site A: £8.50 million. The 95% Confidence Interval for Site A: (4.20, 12.81).
2) Predicted Revenue for Site B: £2.67 million. The 95% Confidence Interval for Site B: (-2.01, 7.35).

Based on our model and historical data, the confidence intervals indicate a range of possible revenues for both Site A and Site B. For Site A, we are 95% confident that the actual revenues are in the range of approximately £4.20 million to £12.81 million. For Site B, however, the confidence interval is somewhat concerning. We are 95% confident that the true revenue falls within the range of approximately -£2.01m to £7.35m. This confidence interval includes values that could potentially be negative, suggesting that there is a significant level of uncertainty in predicting Site B's revenue, with the possibility that it could fall into a negative range. With this in mind, it's prudent to lean towards the assumption that Site A has better revenue prospects, as the true value for Site B could potentially be in the negative range.

Nevertheless, to reduce the standard error and to be more certain about the predictive revenues, we recommend collecting additional data, such as details on sale types, regional demographics, and client profiles, to conduct more conclusive analyses in the future. These insights will enable us to make more informed decisions and refine our predictions. Additionally, our analysis shows that quality of management explains ~75% of the variance in the data and there seems to be a direct relationship between quality of management and revenue. Therefore, we also suggest finding a competent manager for whichever site one decides to invest in.

It is important to consider the uncertainty associated with these predictions, as we are 95% confident that the true revenue for Site A falls within the range of approximately £4.20 million to £12.81 million, and for Site B, it falls within the range of approximately £-2.01 million to £7.35 million.

These confidence intervals account for the inherent variability in our model and provide a more complete picture of the potential revenue outcomes for both sites.

The prediction equation reveals interesting insights into the effects of income on earnings. The negative coefficient for INCOME (-4.3543) suggests that as the local income level increases, revenue tends to decrease, up to a certain point, which might seem counterintuitive and which is evident in the lower performance of site B compared to site A, despite having a higher per capita income. This is because higher-income areas may have different preferences or competitive dynamics.

However, it is essential to consider the nuanced relationship. The positive coefficient for the squared income term (0.1755) suggests a quadratic relationship, indicating that beyond a certain point, income increases tend to have a positive impact on revenue. This underlines the importance of aligning pub strategies, themes, and pricing with the local income context to optimize performance. Moreover, even assuming a higher management quality scale (MAN) for site A, it is important to note that this variable appears as the most significant explanatory factor for revenue variability within our model (as illustrated in Figure 7, Percentage of Variance). Therefore, the crucial importance of prioritizing exceptional management cannot be overstated, as it continues to play a substantial role in the achievement of superior revenue.

Next to the less concerning confidence interval, acquiring site A would also have the advantage of allowing us to improve our forecasting model for the future. We had to exclude COM from our regression due to heteroscedasticity. One reason for this was that we lacked data points for more pubs with less competition. By acquiring site A, which has a low density of competition, we gain more insight into how competition affects revenue streams. Normally one would assume that less competition, as we have in site A, is better than more competition, as we have in site B.

Appendix (List of Figures)

***Figure 1:* Scatter Plots**

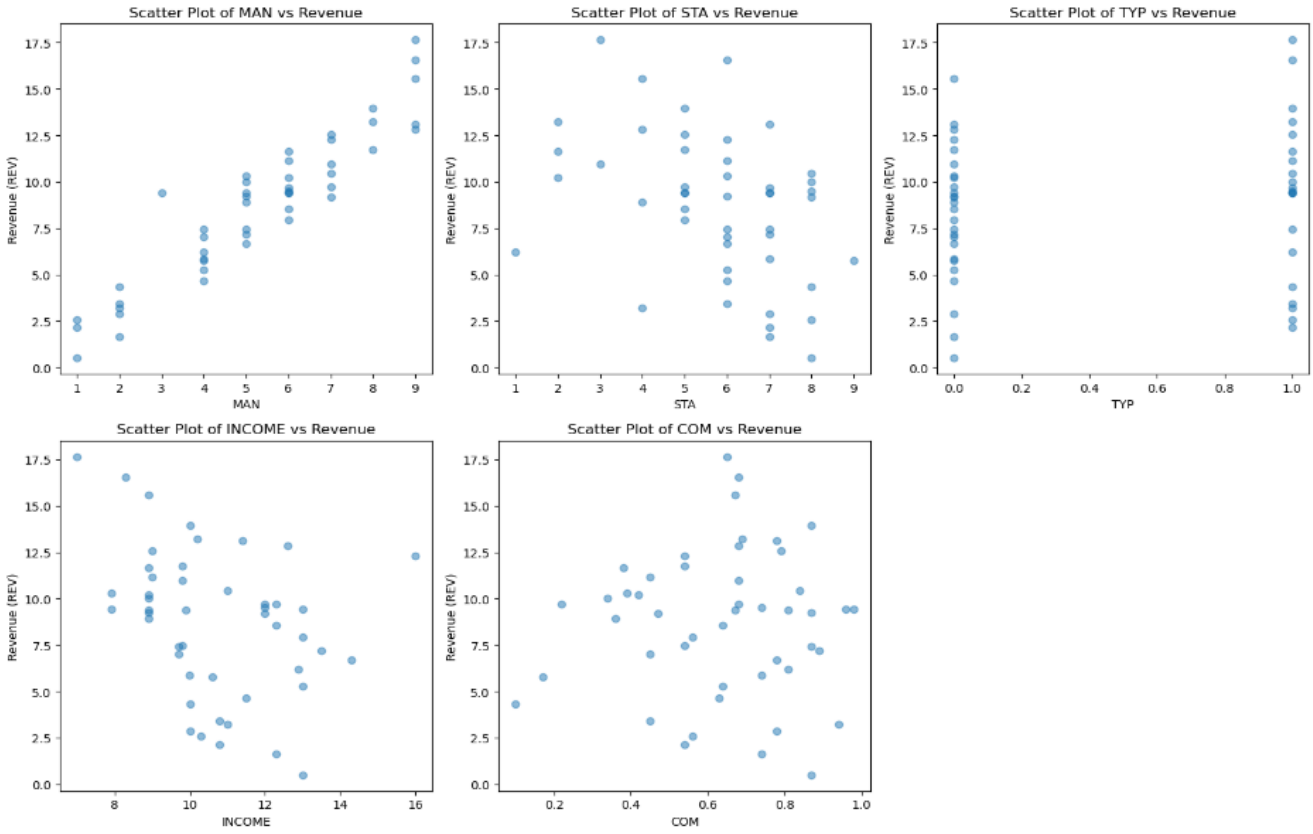


***Figure 2:* 5 Variables**

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.975007
R Square	0.950638
Adjusted F	0.944619
Standard E	0.929751
Observatic	47

ANOVA

	df	*SS*	*MS*	*F*	*ignificance F*
Regressio	5	682.5656	136.5131	157.9213	1.16E-25
Residual	41	35.44193	0.864437		
Total	46	718.0075			

	Coefficient	*tandard En*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	4.635018	1.133558	4.088911	0.000197	2.345751	6.924284
MAN	1.588607	0.065938	24.09241	8.02E-26	1.455443	1.721772
STA	-0.03392	0.080819	-0.41974	0.67687	-0.19714	0.129295
TYP	1.087024	0.289447	3.755517	0.000538	0.502474	1.671575
INCOME	-0.43257	0.078596	-5.50372	2.19E-06	-0.5913	-0.27384
COM	-0.04553	0.671414	-0.06781	0.946267	-1.40148	1.310421

***Figure 3:* 4 Variables**

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.975004
R Square	0.950633
Adjusted F	0.945931
Standard I	0.918668
Observatic	47

ANOVA

	df	*SS*	*MS*	*F*	*ignificance F*
Regressio	4	682.5616	170.6404	202.1925	7.65E-27
Residual	42	35.44591	0.84395		
Total	46	718.0075			

	Coefficient!tandard En		*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	4.618864	1.095031	4.218022	0.000128	2.409002	6.828725
MAN	1.588347	0.065042	24.42035	1.92E-26	1.457087	1.719607
STA	-0.03361	0.079723	-0.42154	0.675507	-0.19449	0.127281
TYP	1.084488	0.2836	3.824009	0.000429	0.512161	1.656816
INCOME	-0.4337	0.075893	-5.7146	1.02E-06	-0.58686	-0.28054

***Figure 4:* 3 Variables**

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.974897
R Square	0.950424
Adjusted F	0.946965
Standard I	0.909841
Observatic	47

ANOVA

	df	*SS*	*MS*	*F*	*ignificance F*
Regressio	3	682.4117	227.4706	274.7856	4.61E-28
Residual	43	35.59588	0.827811		
Total	46	718.0075			

	Coefficient!tandard En		*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	4.39706	0.951106	4.623102	3.44E-05	2.478972	6.315148
MAN	1.598929	0.059426	26.90618	1.56E-28	1.479085	1.718773
TYP	1.090089	0.280567	3.885309	0.000348	0.524272	1.655905
INCOME	-0.43649	0.074877	-5.82944	6.47E-07	-0.5875	-0.28549

Figure 5: **Residual Plots**

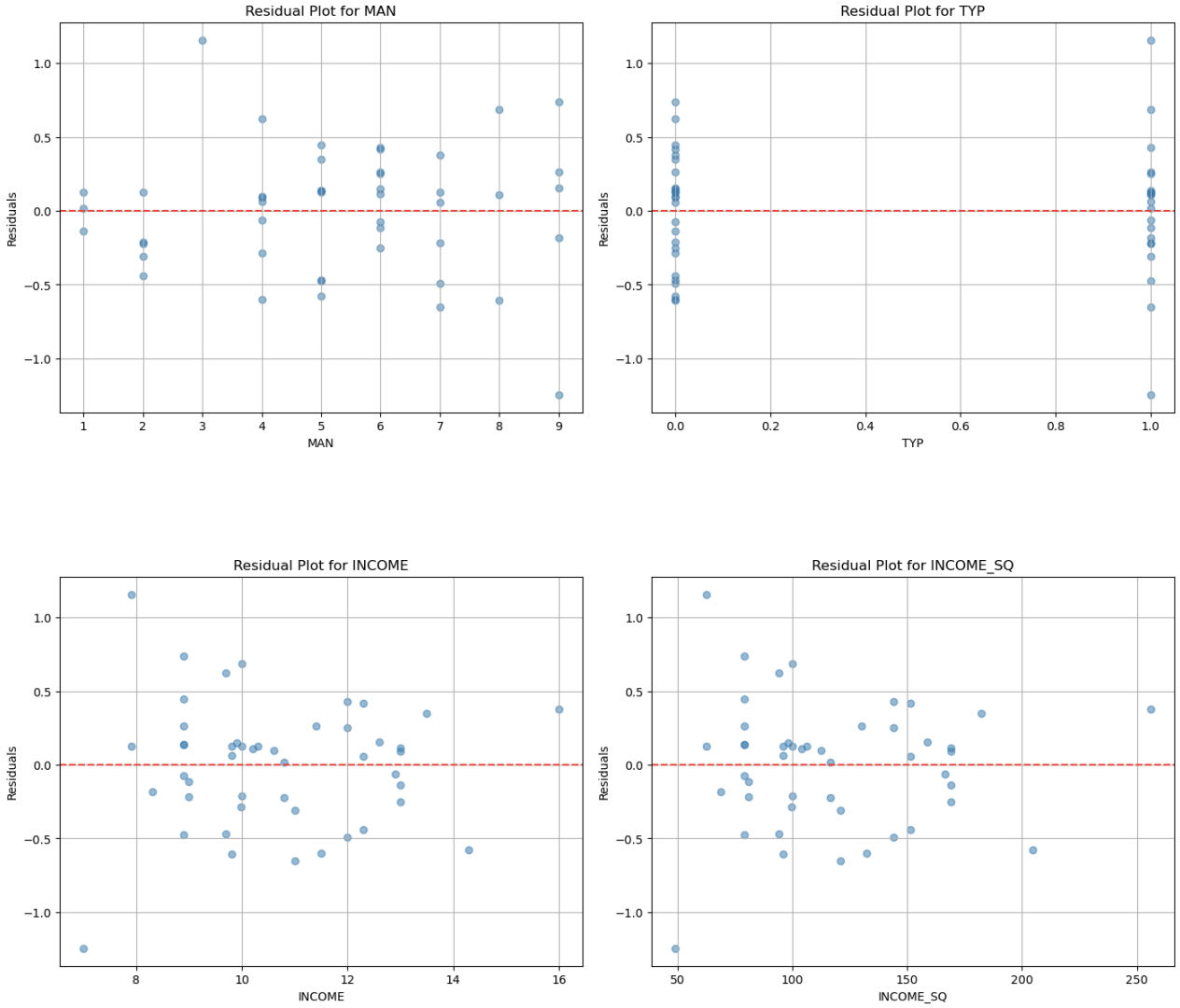


Figure 6: **3 Variables + Quadratic Term**

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0,994
R Square	0,988
Adjusted R S(	0,987
Standard Erro	0,446
Observations	47

ANOVA

	df	*SS*	*MS*	*F*	*ignificance F*
Regression	4	709,6688	177,4172	893,6081	5,03E-40
Residual	42	8,338692	0,19854		
Total	46	718,0075			

	Coefficients	*tandard Err*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*
Intercept	26,0577	1,906429	13,66833	4,66E-17	22,21037	29,90503
MAN	1,5126	0,030021	50,3864	3,41E-39	1,452042	1,573209
TYP	1,1021	0,137406	8,021019	5,23E-10	0,824842	1,379437
INCOME	-4,3543	0,336375	-12,9448	2,98E-16	-5,03314	-3,67548
INCOME_sq	0,1755	0,014981	11,71699	8,06E-15	0,145303	0,20577

Figure 7: **Percentage of Variance Explained by Each Variable**