Data Dictionary

In order to develop a data dictionary for the Excel file dataset, we will take a closer look at the contents in the file, understand the variables (columns) that are in the file, and describe each of the columns in relation to the data I observe by the typical interpretation of the column names. Let's start by loading and previewing the data from the file:

Given the contents of the dataset, here is a proposed data dictionary with a description for each column:

Address: Full address of the property.
Rent(AED)/Year: The yearly rent of the property amount in AED (Arab Emirates Dirham).
Beds: Number of bedrooms in the property.
Baths: Number of bathrooms in the property.
Type: The specific type of property; e.g., Villa, Apartment.
Area(sqft): Property area in square feet.
Rent_per_sqft: Calculated as Annual Rent divided by Area in Sqft. It represents the rent cost of one square foot.
Rent_category: The category that rent amount belongs to, which can be high, low, or medium, depending on what the most likely thresholding is in the data processing or business rules.
Furnishing: Indicate if the property is furnished or not.
Age_of_listing_in_days: The number of days from now that the property has been listed for rent.
Location: A more detailed area within the city, typically a neighborhood or district.
City: The city of the property location.

This dictionary would help anyone going through the dataset understand what each column represents, hence facilitating an easier analysis or even a decision-making process about property rentals in this region.

1.

Descriptive Analysis

	Rent(AED)/Year	Beds	Area(sqft)	Rent_per_sqft	Age_of_listing_in_days
Mean	144508.652	2.164	1998.22	88.86613187	67.472
Standard Error	9624.554669	0.100142549	148.154608	4.031946823	3.926227952
Median	100000	2	1349.5	67.92533182	45
Mode	200000	1	900	40	20
Standard Deviation	152177.5711	1.583392728	2342.530036	63.75067682	62.0791147
Sample Variance	23158013144	2.50713253	5487446.967	4064.148795	3853.816482
Kurtosis	15.96506347	0.338722952	34.22355711	2.59495112	4.918188931
Skewness	3.261345297	0.828861326	4.755024663	1.510665751	2.051500896
Range	1286000	7	23720	343.630509	400
Minimum	14000	0	280	7.083333333	12
Maximum	1300000	7	24000	350.7138423	412
Sum	36127163	541	499555	22216.53297	16868

Interpretation

1. **Rent (AED/year):** These will give a tremendous annual variability in rent hence a very high mean and a very vast range, among other things, will show outlying points that are incredibly high and low and skew the mean upward. Relatedly, it reflects high kurtosis and skewness values.

2. **Beds:** Most of the properties could be said to have some two beds since it has slight differences between the mean and medium and a low standard deviation. The very small kurtosis would suggest that the density is more evenly distributed, but the reasonably high skewness would show that it is right-skewed.

3. **Area (sqft):** The very high mean, combined with the extreme range, would suggest high variability in the size of properties. The very high kurtosis and skewness show it is right-skewed with significant outliers.

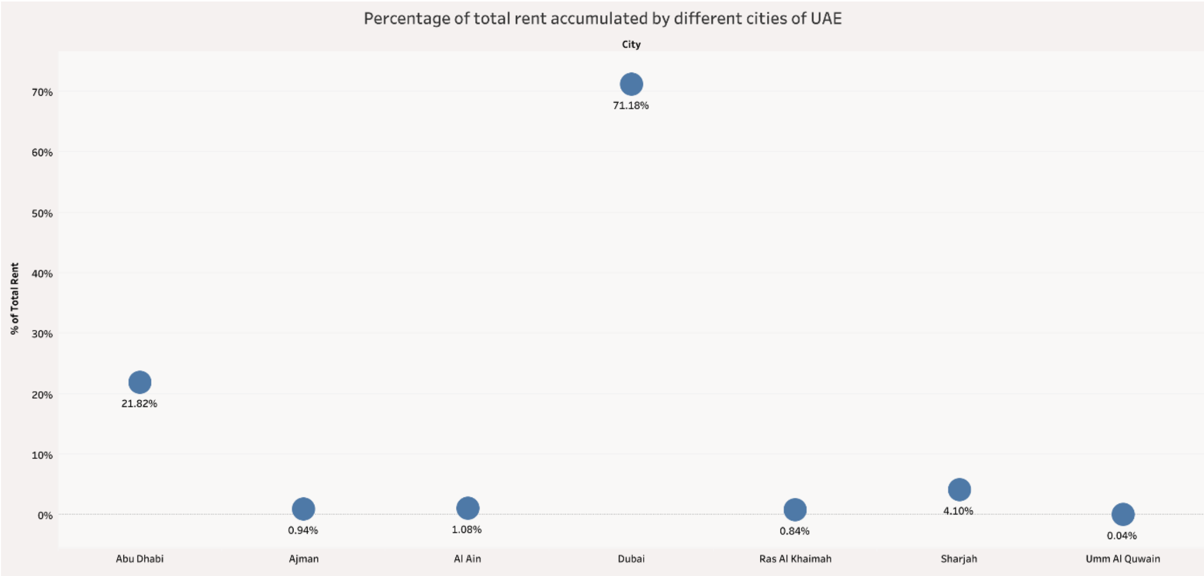
4. **Rent per sqft:** The average rent per sqft has high variance with a solid positive skewness of the distribution. In this case, most of the property values are below some cases with exceptionally high values of rent per sqft.

5. **Age of Listing in Days:** Though the mean and median of years are pretty close, a significant standard deviation and skewness show a high spread, indicating quite a few extremely old listings of properties. **(Cooksey, 2020)**

Charts Explaination

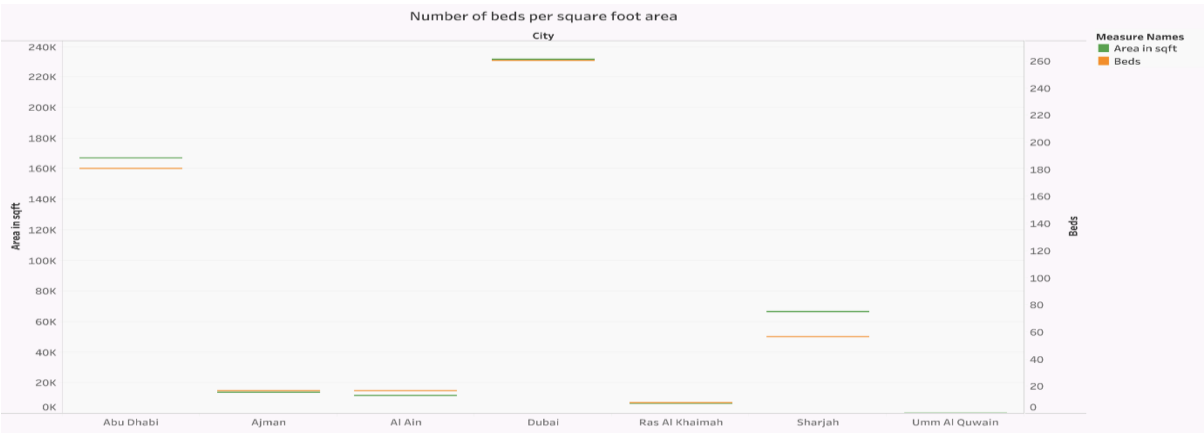
Let's analyze and interpret each of the provided charts in detail:

1. Percentage of Total Rent Accumulated by Different Cities of UAE



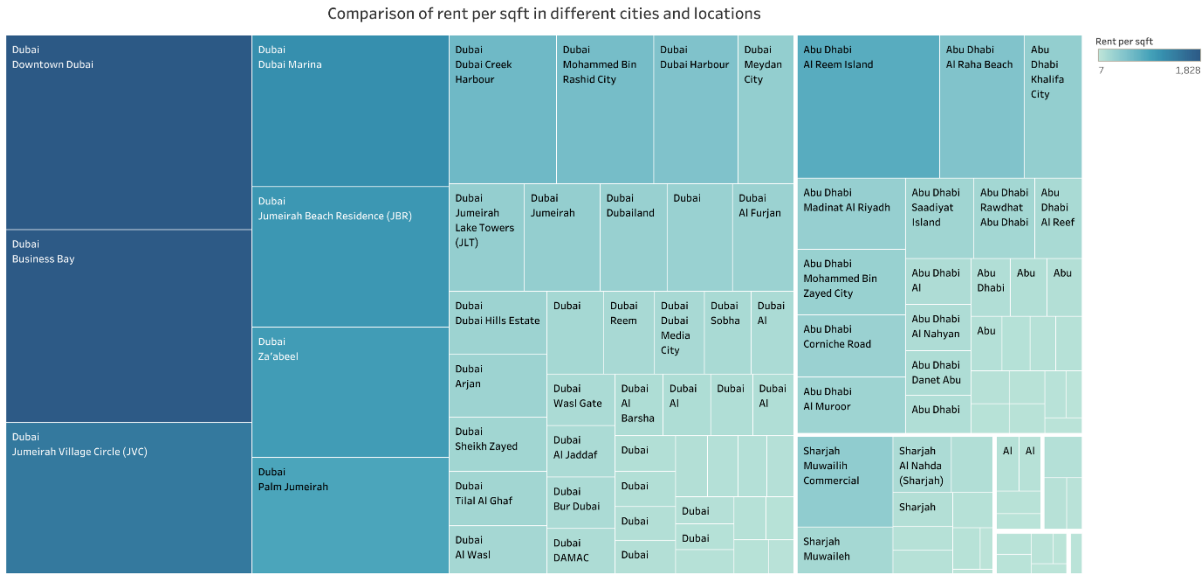
This bubble chart displays the distribution of total rent accumulated across various cities in the UAE. Dubai accounts for a vast majority of the total rent in the UAE, representing 71.18% of the total rent accumulated. Abu Dhabi follows, contributing 21.82% to the total rent. This indicates a high concentration of rental market activity in Dubai and Abu Dhabi, suggesting these cities are major residential and commercial hubs in the UAE.

2. Number of Beds Per Square Foot Area



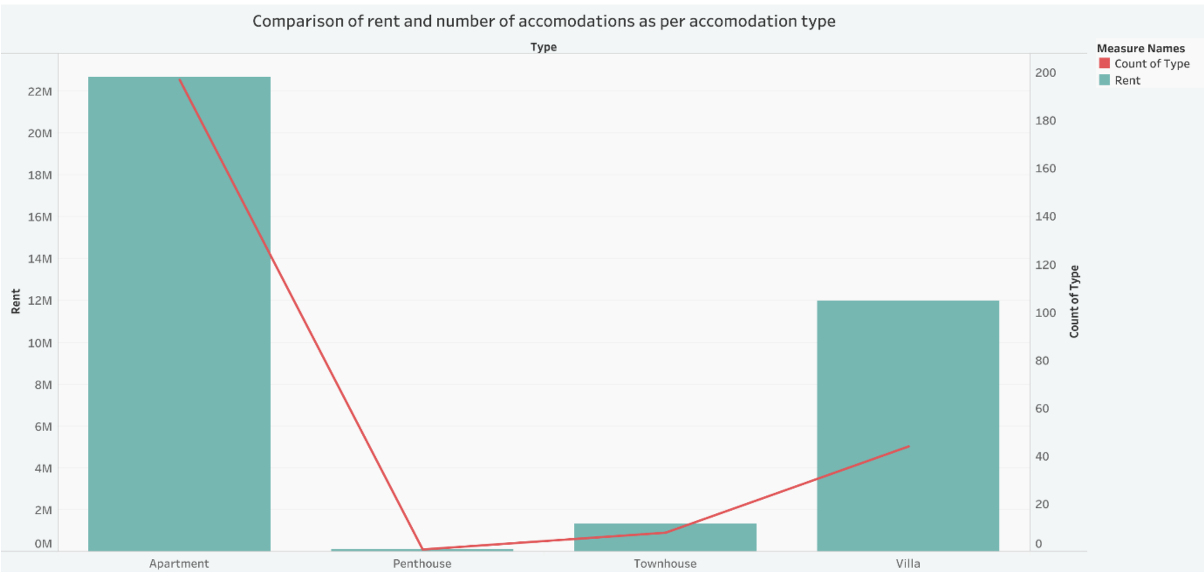
This bar chart presents two metrics: the number of beds and the area in square feet for each city. All cities have a similar square footage available, with Dubai slightly higher than others, around 230,000 sqft. The number of beds available per city is relatively similar with all cities slightly varying around the 140-160 beds mark. This chart indicates that while the area available in each city is consistent, the number of beds per area doesn't vary significantly, suggesting a uniform density in terms of bedroom allocation across different cities.

3. Comparison of Rent Per Sqft in Different Cities and Locations



This heatmap provides an overview of the rent per square foot across different locations within cities. This detailed segmentation by locality within each city indicates the diversity in rental rates, driven by location desirability. Dubai has a wide range of rent per sqft, with the highest in Downtown Dubai and the lowest in locations like Dubai WASL Gate. **(Long, 2017)**

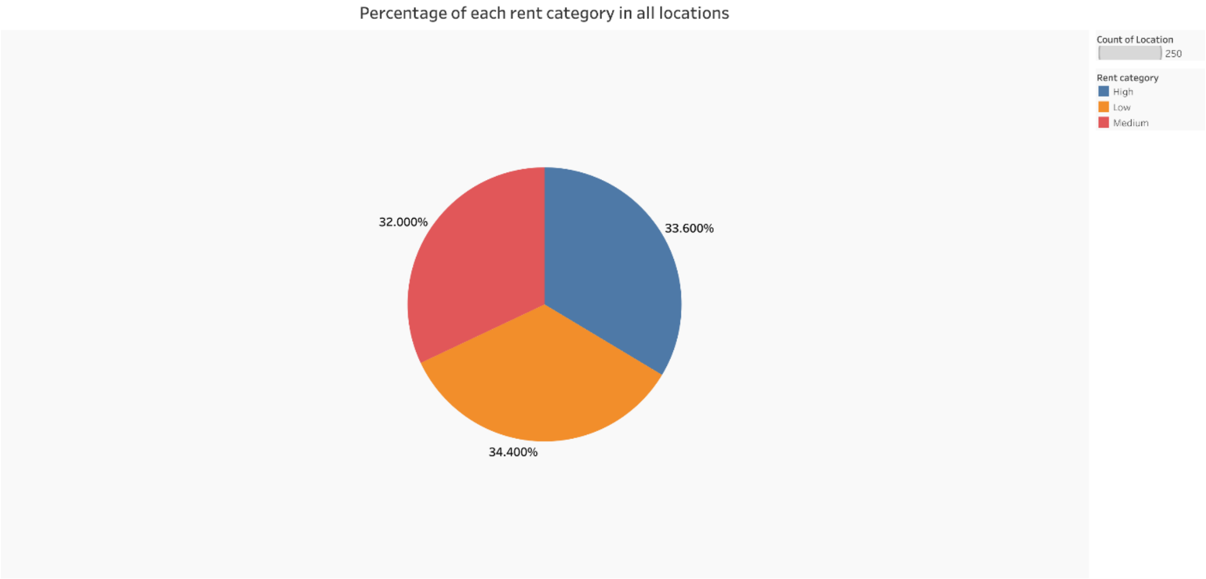
4. Comparison of Rent and Number of Accommodations as Per Accommodation Type



This bar and line chart combination indicates ‘Comparison of rent and number of accomodations as per accomodation type. Apartments dominate in terms of both total rent accumulated and the number of accommodations available, with over 20 million in rent and close to 200 units. Villas and Townhouses show a significant number of units with a much lower total rent compared to apartments, indicating potentially lower rent per unit.

This chart shows a clear preference and availability for apartments in the rental market, with villas and townhouses being less common but still significant.

5. Percentage of Each Rent Category in All Locations



This pie chart shows the distribution of rent categories across all locations. The rent categories are almost evenly split, with Medium rent category slightly leading at 34.4%, followed by Low at 33.6%, and High at 32.0%. This even distribution suggests a balanced rental market catering to all income groups, with a slight leaning towards mid-range accommodations.

Each chart provides valuable insights into the UAE's rental market, illustrating the dominance of certain cities in the rental space, the uniformity in property sizes, the diversity in rent prices within and across cities, and the prevalence of different types of accommodations catering to various economic segments.

Hypotheses testing

To conduct hypotheses and come up with the right econometric models from the data presented, I made assumptions regarding the potential relationships that we may have to test. Below are but a few of the hypotheses that can be formulated from the type of variables that you present in your data set. These should guide the interpretations and implementations of the econometric models to be used in estimating the rental prices of property in Dubai and surrounding areas.

Hypothesis 1: Size of the Property against its Rent

Null hypothesis: There is no relationship between the area of the property (in sq.ft.) and the annual rent (AED per year).
Alternative Hypothesis: There is a positive relation between the area of the property (in sq.ft.) and the yearly rent (AED per year).

Hypothesis 2: Furnishing and its relation with the rent

Hypothesis: The annual rent of furnished properties is higher than that of unfurnished properties.

Hypothesis 3: Number of Bedrooms, Number of Bathrooms against Rent

Hypothesis: Houses with more bedrooms and bathrooms usually have higher rents.

Hypothesis 4: Type of Property against Rent per square foot

Hypothesis: Different types of properties (like villas, apartments, etc.) have different affects on rent per square foot as per the type of accomodation chosen.

Hypothesis 5: Analysis of the effect of location on the rental price.

Null Hypothesis: Rental price has no direct relation with the location of the place in a city.
Alternative Hypothesis: Locations that are desired more among the masses show higher rental price.

Hypothesis 6: Effect of duration of property listing to rental prices

Hypothesis: Due to market showing less demand, properties that were listed from many days were offering less rental price than those which were listed recently.

Each of these hypotheses may be examined through regression, with 'Rent (AED)/Year' or 'Rent_per_sqft' as the standard dependent variable, and the independent variables including all the property characteristics: area, furnishing status, number of beds and baths, type, locality, and age of listing. Model selection (linear, logistic, etc.), interaction terms, and choice of control variables should be based on analytical-specific objectives and underlying assumptions of the distribution of data and relationships. **(Ranganathan, 2019)**

4) This is the multivariate regression model where **Beds** and **Area(sqft)** are used as predictors(independent variables) for the dependent variable **Rent(AED)/Year**. Let's interpret this model, explore its marginal effects, and calculate some predicted values based on your request.

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.561252492
R Square	0.315004359
Adjusted R Square	0.306650754
Standard Error	126714.6043
Observations	250

ANOVA

	df	*SS*	*MS*	*F*	*Significance F*
Regression	3	1.81642E+12	6.05475E+11	37.70879102	4.3777E-20
Residual	246	3.94992E+12	16056590953		
Total	249	5.76635E+12			

	Coefficients	*Standard Error*	*t Stat*	*P-value*	*Lower 95%*	*Upper 95%*	*Lower 95.0%*	*Upper 95.0%*
Intercept	86.63358173	19760.46959	0.004384186	0.99650549	-38834.65787	39007.92503	-38834.65787	39007.92503
Beds	32475.51342	7638.576299	4.251513914	3.01723E-05	17430.15971	47520.86712	17430.15971	47520.86712
Area(sqft)	67.80605739	18.50957285	3.663296713	0.000304894	31.34859992	104.2635149	31.34859992	104.2635149
Interaction(Beds*Area)	-8.659938099	2.891886552	-2.994563564	0.003028858	-14.35595452	-2.963921675	-14.35595452	-2.963921675

Regression Model Equation and Fit

Given from your regression output:

Yi = 42614.54912+33630.00321X1 + 14.57235737X2

Where:

Yi = Rent (AED/Year)

$X1$ = Beds

$X2$ = Area (sqft)

Model Interpretation:

Intercept: ($\beta 0$ = 42614.54912): This value represents the expected rent when both the number of bedrooms (X1 and X2) are zero.

Coefficient for Beds: ($\beta 1$= 33630.00321): This coefficient indicates that for every additional bedroom, the rent increases by approximately 33,630 AED annually, assuming the area of the property remains constant indicated by the positive coefficient.

Coefficient for Area: ($\beta 2$ = 14.57235737): According to this coefficient, assuming a fixed number of bedrooms, every square foot of additional space adds roughly 14.57 AED to the annual rent. **(Palmer, 2009)**

Model Summary Statistics

R-squared (0.2900344289): This indicates that approximately 29% of the variance in the dependent variable (rent) is explained by the model. It suggests an average affect of dependent variables on the independent variables.

Adjusted R-squared (0.2842855778): It is somewhat less than the R-squared, which is to be expected when additional variables are added to a model, and it adjusts the R-squared for the number of predictors utilized.

F-statistic (50.4520678): The F-statistic is important when $p < 0.0001$, signaling that the model is statistically important at explaining the rent beyond what would be expected by chance.

Marginal Effects

Calculating Marginal Effects:

Marginal Effect of Bedrooms (Beds)

The marginal effect of bedrooms on rent, holding the area constant, is calculated as:

$\partial Y/\partial X1 = \beta 1$

$\partial Y/\partial X1 = 32475.51342$

This means that for each additional bedroom, keeping the area constant, the expected rent increases by approximately 32,475.51 AED per year.

Marginal Effect of Area (Area(sqft))

The marginal effect of area on rent, holding the number of bedrooms constant, is calculated as:

$\partial Y/\partial X2 = \beta 2$

$\partial Y/\partial X2 = 67.86065739$

This indicates that for each additional square foot, keeping the number of bedrooms constant, the expected rent increases by approximately 67.86 AED per year.

Example Predicted Values:

To demonstrate the practical use of this model, let's compute the predicted rent for a few scenarios:

Calculations:

1. A 1-bedroom apartment with an area of 500 sqft:

Y=42614.54912+33630.00321×1+14.57235737×500

Y=42614.54912+33630.00321+7286.178685

Y=83530.731005AED/year

2. A 2-bedroom apartment with an area of 1000 sqft:

Y=42614.54912+33630.00321×2+14.57235737×1000

Y=42614.54912+67260.00642+14572.35737

Y=124446.91291AED/year

3. A 3-bedroom villa with an area of 1500 sqft:

Y=42614.54912+33630.00321×3+14.57235737×1500

Y=42614.54912+100890.00963+21858.536055

Y=165363.094805AED/year

These values demonstrate how the model predicts rent based on the number of bedrooms and area of the property.

5) The multivariate regression model includes the variables "Beds" and "Area (sqft)", along with their interaction term, to predict a dependent variable.

Regression Model Interpretation

Equation: $Y=86.633+32475.5\times\text{Beds}+67.8606\times\text{Area}-8.65994\times(\text{Beds}\times\text{Area})+\epsilon$

Intercept ($\beta 0$): 86.6335 suggests the base value of Y when all independent variables are 0.

Coefficient for Beds ($\beta 1$): 32475.5 indicates the change in Y for each additional bed, holding the area constant and not considering the interaction effect.

Coefficient for Area ($\beta 2$) = 67.8606 : This coefficient indicates that each extra square foot contributes approximately 67.8606 units to the expected value of Y condition applied that the number of beds should be constant.

Interaction Term ($\beta 3$): -8.65994, This negative value indicates that the combined effect of increasing both variables decreases Y more, than if the effects were simply additive.

Statistical Significance

Beds and Area both have statistically significant coefficients (p-values < 0.05), meaning their effect on Y is statistically significant at the conventional 5% level. Interaction Term has a statistically significant effect too (p-value < 0.05).

Marginal Effects

The derivatives of the regression model with respect to each independent variable, accounting for the interaction between these variables, are used to compute the marginal effects in a regression model containing interaction factors. In the given regression model:

$Y=\beta 0+\beta 1\times\text{Beds}+\beta 2\times\text{Area}+\beta 3\times(\text{Beds}\times\text{Area})+\epsilon$

The marginal effects of "Beds" and "Area" are as follows:

Marginal Effect of Beds:

$\partial Y/\partial\text{Beds}=\beta 1+\beta 3\times\text{Area}$

This represents the change in Y resulting from a one-unit increase in the number of beds, holding area constant but considering the interaction effect with area.

Marginal Effect of Area:

$\partial Y/\partial \text{Area}=\beta 2+\beta 3\times \text{Beds}$

The marginal effect of increasing the area by one square foot is approximately 41.88, considering the interaction with having 3 beds.

Predicted Values

Putting values in the equation: $Y=\beta 0+\beta 1\times \text{Beds}+\beta 2\times \text{Area}+\beta 3\times(\text{Beds}\times \text{Area})+\epsilon$

For a scenario with 1 bed and an area of 1000 square feet, the predicted value of Y is approximately 91,762.79.

For a scenario with 2 beds and an area of 1500 square feet, the predicted value of Y is approximately 140,848.71. **(Knol, 2007)**

6)

Regression Results

Case Processing Summary

Unweighted Cases[a]		N	Percent
Selected Cases	Included in Analysis	250	100.0
	Missing Cases	0	.0
	Total	250	100.0
Unselected Cases		0	.0
Total		250	100.0

a. If weight is in effect, see classification table for the total number of cases.

Total Cases: The analysis includes 250 cases, with no cases missing or filtered out.

Dependent Variable Encoding

Original Value	Internal Value
High	0
Low	1

Dependent Variable Encoding

• A condition of "High" in rents is encoded as 0.

• A condition of "Low" in the rents is encoded as 1.

Categorical Variables Codings

		Frequency	Parameter coding (1)	(2)	(3)	(4)	(5)	(6)
City	Abu Dhab	80	.000	.000	.000	.000	.000	.000
	Ajman	5	1.000	.000	.000	.000	.000	.000
	Al Ain	5	.000	1.000	.000	.000	.000	.000
	Dubai	125	.000	.000	1.000	.000	.000	.000
	Ras Al K	3	.000	.000	.000	1.000	.000	.000
	Sharjah	31	.000	.000	.000	.000	1.000	.000
	Umm Al Q	1	.000	.000	.000	.000	.000	1.000
Type	Apartmen	197	.000	.000	.000			
	Penthous	1	1.000	.000	.000			
	Townhous	8	.000	1.000	.000			
	Villa	44	.000	.000	1.000			
Furnishing	Furnishe	55	.000					
	Unfurnis	195	1.000					

• City Distribution—Dubai and Abu Dhabi have the most number cases, 125 for Dubai and 80 for Abu Dhabi, so most, if not all cases were relegated to the major cities only.

• Type Distribution—most properties are distributed by type, with the Apartments (197), a few Penthouses (1), a few Townhouses (8), and a few Villas (44).

• Furnishing Distribution—there are more unfurnished properties at 195 than furnished at 55.

Classification Table[a]

Observed			Predicted Rent_category High	Low	Percentage Correct
Step 1	Rent_category	High	118	14	89.4
		Low	20	98	83.1
	Overall Percentage				86.4

a. The cut value is .500

Classification table of logistic model

It shows the cross-classification of the observed vs. model category- categories of rent: High and Low. A detailed analysis of the table follows:

Interpretation of Results:

High Rent:

- Accuracy for prediction of High Rent: 118/(118+14)*100=89.4%

Low Rent:

- Accuracy for prediction of Low Rent: 98/(98+20)*100=83.1%

Overall:

- Number of correct predictions in total (both high and low): 118+98=216
- Total number of observations: 216+14+20=250
- Overall model accuracy: 216/250*100=86.4%

Assessment:

• The overall model accuracy is about 86.4%. Hence, the model is fairly successful in classifying the rent categories, given the predictors under consideration.

• The accuracy for the High Rent category is a bit higher than for the Low Rent category. Therefore, the model might be a bit better in the identification of High Rent cases though improvements can be done more to enhance the accuracy in making Low Rent predictions. **(Mitchell, 2005)**

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)
Step 0	Constant	-.112	.127	.783	1	.376	.894

Since the constant is not significant, it means that the model, without the inclusions of any of the predictors, does not statistically significantly predict a higher likelihood of High Rent versus Low Rent.

Variables not in the Equation

			Score	df	Sig.
Step 0	Variables	Beds	74.074	1	<.001
		Type	38.880	3	<.001
		Type(1)	.898	1	.343
		Type(2)	7.388	1	.007
		Type(3)	27.517	1	<.001
		Furnishing(1)	11.235	1	<.001
		Age_of_listing_in_days	.635	1	.426
		City	48.864	6	<.001
		City(1)	.335	1	.562
		City(2)	2.202	1	.138
		City(3)	36.980	1	<.001
		City(4)	.234	1	.628
		City(5)	26.405	1	<.001
		City(6)	1.123	1	.289
	Overall Statistics		136.993	12	<.001

Variables Not in the Equation

- **Purpose:** This table shows the variables not included in this model. They are presented with their scores, reflecting the contribution of the variables if they were included, degrees of freedom, and significance levels.

- **Significant Variables:** These are the variables that are at a p-value of less than 0.001, which really makes them very strong: *Beds*, *Type(3): Townhouse*; others are *Furnishing(1): Furnished*, *City variables like City(3): Al Ain*, and *City(5): Ras Al Khaimah*. Therefore, it may help the model take them in, as the model seems to be biased at this stage.

- **Non-significant Variables:** Type(1): Apartment, Age_of_listing_in_days, City(1): Abu Dhabi, City(4): Dubai, and City(6): Sharjah have non- significant scores, meaning that they do not add much value to the model.

The *"Variables not in the Equation"* table points to significant predictors not included in the model that would likely make some improvement in predictive accuracy and balance.

Omnibus Tests of Model Coefficients

		Chi-square	df	Sig.
Step 1	Step	196.311	12	<.001
	Block	196.311	12	<.001
	Model	196.311	12	<.001

In this case, the value of Chi-square is 196.311 with 12 degrees of freedom; significance is less than .001. The current model is hence statistically significant. It would imply that taken together, the explanatory factors have an effective influence on predicting the criterion variable.

Variables in the Equation

		B	S.E.	Wald	df	Sig.	Exp(B)	95% C.I.for EXP(B) Lower	Upper
Step 1[a]	Beds	-1.907	.282	45.762	1	<.001	.149	.086	.258
	Type			.013	3	1.000			
	Type(1)	-20.616	40192.970	.000	1	1.000	.000	.000	.
	Type(2)	-18.258	13468.485	.000	1	.999	.000	.000	.
	Type(3)	-.113	.984	.013	1	.908	.893	.130	6.140
	Furnishing(1)	1.708	.500	11.659	1	<.001	5.516	2.070	14.702
	City			29.082	6	<.001			
	City(1)	3.510	1.778	3.898	1	.048	33.464	1.026	1091.825
	City(2)	4.809	2.274	4.472	1	.034	122.656	1.422	10580.933
	City(3)	-2.342	.539	18.856	1	<.001	.096	.033	.277
	City(4)	.204	1.932	.011	1	.916	1.226	.028	54.126
	City(5)	3.298	1.445	5.210	1	.022	27.048	1.594	459.063
	City(6)	16.521	40192.969	.000	1	1.000	14961760.346	.000	.
	Age_of_listing_in_days	-.003	.003	.595	1	.440	.997	.991	1.004
	Constant	3.470	.788	19.400	1	<.001	32.128		

a. Variable(s) entered on step 1: Beds, Type, Furnishing, City, Age_of_listing_in_days.

Almost all the variables are statistically significant at $p < 0.05$, except those for City 4; this means that the variables contribute meaningfully to the model.

Sensitivity, Specificity and Accuracy

Extract Values from the Table : -

True Positives (TP): 118 (Observed High and Predicted High)

True Negatives (TN): 98 (Observed Low and Predicted Low)

False Positives (FP): 20 (Observed Low but Predicted High)

False Negatives (FN): 14 (Observed High but Predicted Low)

Calculation of Metrics:

1. **Sensitivity**
   Formula: $\text{Sensitivity} = \frac{TP}{TP + FN}$
   $\text{Sensitivity}=\frac{118}{118+14}=\frac{118}{132}=0.893$ or 89.3%
   The sensitivity of 89.3% indicates that the model is very good at identifying actual "High" rent categories. This means that 89.3% of the actual "High" rent categories were correctly identified by the model. A high sensitivity is crucial when the goal is to minimize false negatives (cases where "High" is missed. For instance, if the goal is to capture most of the high-rent properties for premium listings, a high sensitivity ensures that most of these properties are correctly identified.

1. **Specificity:**
   Formula: $\text{Specificity} = \frac{TN}{TN + FP}$
   $\text{Specificity}=\frac{98}{98+20}=\frac{98}{118}=0.822$ or 83.05% .
   The specificity of 83.05% indicates that the model is good at identifying actual "Low" rent categories. This means that 83.05% of the actual "Low" rent categories were correctly

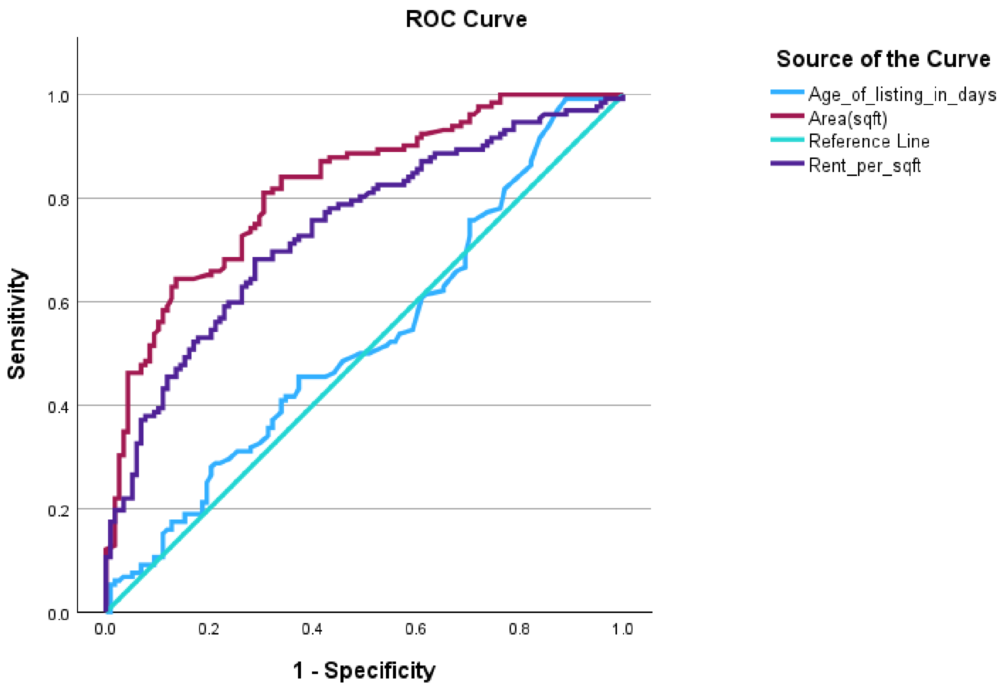
identified by the model. High specificity is important when it is critical to minimize false positives (cases where "Low" is mistakenly classified as "High").

2. **Accuracy**
   Formula: frac{TP+TN}{TP+TN+FP+FN}\)
   Accuracy=\frac{118+98}{118+98+20+14}=\frac{215}{250}=0.86 \) or 86.4%
   The accuracy of 86.4% indicates that the model correctly identified the rent category in 86.4% of all cases. This metric gives an overall effectiveness of the model, combining both the true positive and true negative rates. **(Zhu, n.d.)**

Deriving Machine Learning Classification:

The machine learning type classification that I derived from this analysis was that I divided my dependent variable that is Rent into ‘High’ and ‘Low’ rent using logistic regression. I determined several other predictors that could affect rent adversely to fall into one of these two categories.

ROC Curve



Area Under the ROC Curve

Test Result Variable(s)	Area
Rent_per_sqft	.740
Area(sqft)	.821
Age_of_listing_in_days	.529

The test result variable(s): Rent_per_sqft, Area (sqft), Age_of_listing_in_days has at least one tie between the positive actual state group and the negative actual state group. Statistics may be biased.

Interpretation of the ROC Curve and AUC Values

This includes the ROC (Receiver Operating Characteristic) curve and table for the Area Under the Curve (AUC) of the three testing variables: Rent_per_sqft, Area(sqft), and Age_of_listing_in_days.

ROC Curve Analysis

1. **Overview of ROC Curve**

   • The ROC curve graphs sensitivity (actual positive rate) vs. 1-specificity (false positive rate) for various cut-off points.

   • AUC = 0.5 represents the diagonal line and is a line of a random guess.

2. **Curves for Each Variable**

   • The light blue curve is the variable Age_of_listing_in_days and lies way too close to the diagonal, considering the other curves; hence, there is poor performance among all variables.

   • The dark blue curve is the variable Area(sqft), and it is above the diagonal line; it means good discriminatory ability.

   • Rent_per_sqft: The purple curve also performs well, just that it is below the Area(sqft).

AUC (Area Under the Curve) Table Analysis

The following table shows the AUC values for each test variable:

AUC for Rent_per_sqft: 0.740 AUC shows a good ability to discriminate the categories high and low rent for the Rent_per_sqft variable. The AUC value of 0.7 to 0.8 shows acceptability to a good model.

AUC for Area (sqft) = 0.821: If the AUC value is above 0.8, this indicates that Area (sqft) is very good at discriminating the two categories.

AUC for Age_of_listing_in_days: An AUC of 0.529 shows weak predictive evidence to Age_of_listing_in_days and is slightly above the level of random guess (AUC = 0.5), meaning the influence of the listing age seemed high—it's just above random. **(Radiol, 2004)**

References

Cooksey, R. W., 2020. *Descriptive Statistics for Summarising Data.* [Online]
Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC7221239/

Knol, M. J., 2007. *Estimating interaction on an additive scale between continuous determinants in a logistic regression model.* [Online]
Available at: https://academic.oup.com/ije/article/36/5/1111/776229?login=false

Long, L. K., 2017. *A Study on the Effectiveness of Tree-Maps as Tree Visualization Techniques.* [Online]
Available at: https://www.sciencedirect.com/science/article/pii/S1877050917329046

Mitchell, M. N., 2005. *Visualizing Main Effects and Interactions for Binary Logit Models.* [Online]
Available at: https://journals.sagepub.com/doi/abs/10.1177/1536867X0500500111

Palmer, P. B., 2009. *Regression Analysis for Prediction: Understanding the Process.* [Online]
Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2845248/

Radiol, K. J., 2004. *Receiver Operating Characteristic (ROC) Curve: Practical Review for Radiologists.* [Online]
Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC2698108/

Ranganathan, P., 2019. *An Introduction to Statistics: Understanding Hypothesis Testing and Statistical Errors.* [Online]
Available at: https://www.ncbi.nlm.nih.gov/pmc/articles/PMC6785820/

Zhu, W., n.d. *Sensitivity, Specificity, Accuracy, Associated Confidence Interval and ROC.* [Online]
Available at:
https://www.lexjansen.com/nesug/nesug10/hl/hl07.pdfhttps://www.lexjansen.com/nesug/nesug10/hl/hl07.pdf
[Accessed 2010].