The "*Commuter Tax*":
Quantifying the Relationship Between Rental Price and CBD Accessibility in Brisbane

April Williams

---

1. Executive Summary

This report investigates the financial trade-off between housing costs and CBD accessibility within the Brisbane metropolitan area. By analysing geospatial data from over 6,000 rental listings, the study aims to quantify the "Commuter Tax" - the premium paid for proximity to the city centre.

The analysis utilises the Haversine formula to calculate precise distances for every listing and applies linear regression to model the price trajectory.

2. Visualising the Market

Figure 1 presents a geospatial heatmap of the Brisbane rental market, derived from over 6,000 active listings. The visualisation utilises a colour gradient to represent price density, where yellow indicates premium pricing tiers and purple represents more affordable market segments.

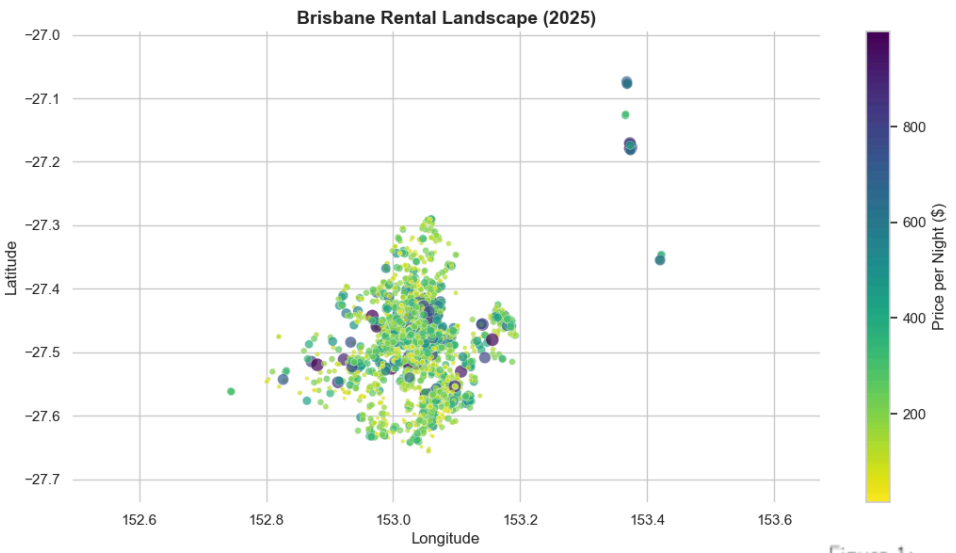

Figure 1:
Geospatial visualisation of the Brisbane rental market

The map illustrates a distinct concentric value structure. The highest concentrations of premium rentals are tightly clustered around the Brisbane River and CBD. As distance from this central spine increases, the market transitions into lower price brackets, visually suggesting a strong correlation between geographic location and rental cost.

3. The “Commuter Tax” Analysis

To quantify the relationship observed in the geospatial analysis, a linear regression model was applied to the dataset (**Figure 2**). This scatter plot maps individual listing prices against their precise Haversine distance from the CBD.

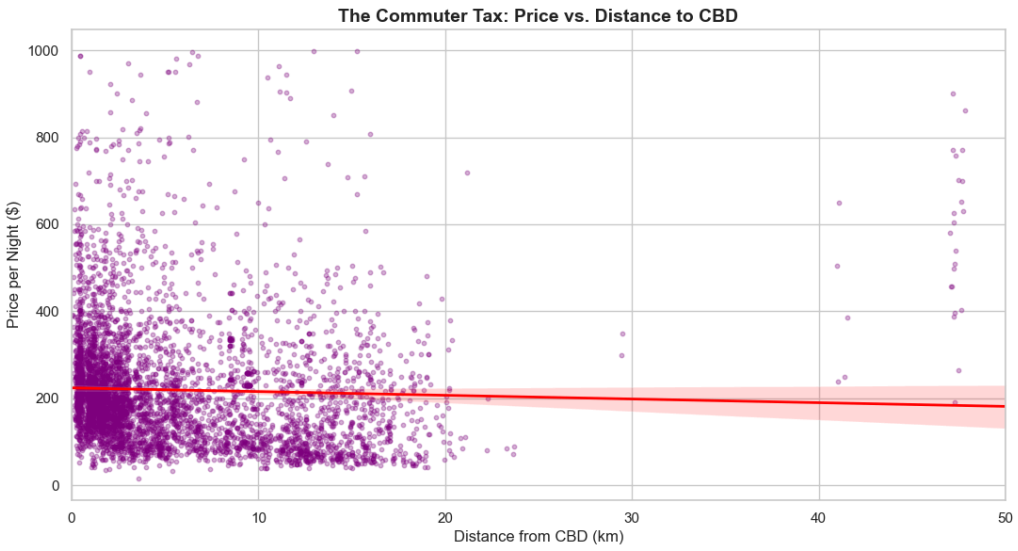

Figure 2:
Scatter plot and linear regression model illustrating the negative correlation between distance from the CBD (km) and nightly rental price ($)

The red trend line demonstrates a clear negative correlation: as distance from the city centre increases, the baseline price decreases. This validates the "Commuter Tax" hypothesis - that tenants pay a measurable financial premium for proximity.

However, the data also highlights significant market complexity. While the general trajectory is downward, distinct clusters of high-value properties remain visible at the 30km mark. These outliers indicate that while distance is a primary driver of price, it is not the sole

determinant, with coastal or lifestyle amenities creating secondary pockets of value that defy the linear trend.

4. Key Business Insights

4.1. Validation of the "Commuter Tax"

The regression analysis (**Figure 2**) confirms a negative correlation between distance and price. As illustrated by the red trend line, rental costs steadily decrease as distance from the CBD increases, validating the hypothesis that tenants pay a measurable premium for proximity to the city.

4.2. The "Coastal" Anomaly

The scatter plot reveals a significant cluster of high-price listings approximately 30km from the CBD (visible as the purple and blue spikes on the far right of Figure 2). This anomaly likely corresponds to the Redcliffe Peninsula. It indicates that "Blue Space" (ocean proximity) commands a market premium comparable to inner-city convenience, breaking the linear rule of "further = cheaper."

4.3. The "Value Zone"

The data identifies a clear "sweet spot" for value-conscious renters between the 15km and 20km marks. In this band, the baseline price (the density of lower points) drops significantly, offering the optimal trade-off between commute time and budget.

Appendix A: Technical Methodology

1. Data Source

This analysis utilises a dataset from Inside Airbnb, accessed in December 2025. The dataset comprises detailed listings for the Brisbane region, including geospatial coordinates (latitude, longitude) and pricing information (price). To ensure the analysis reflected the typical residential market, the following cleaning protocols were applied:

- Price Normalisation: Currency symbols and formatting were removed to convert pricing into numerical float values.
- Outlier Removal: Listings with a nightly price exceeding $1,000 were excluded. This threshold was selected to prevent luxury outliers (e.g., event venues, mansions) from

skewing the regression model, ensuring the findings represent the accessible rental market.

- Missing Values: Rows lacking critical geospatial or pricing data were removed to maintain dataset integrity.

2. Feature Engineering

To test the "Commuter Tax" hypothesis, a new feature, dist_cbd_km, was engineered. This variable represents the direct distance between each listing and the Brisbane CBD (defined as coordinates: -27.4698, 153.0251).

- Calculation Method: Distances were calculated using the Haversine Formula, which accounts for the curvature of the Earth, providing a more accurate measurement than simple Euclidean distance for geospatial data.

3. Statistical Modelling & Visualisation

- Geospatial Analysis: The Seaborn and Matplotlib libraries were used to generate a coordinate-based scatter plot, visualising price density across the metropolitan area.
- Regression Analysis: A linear regression model (sns.regplot) was fitted to the Price vs. Distance data to identify the correlation coefficient and visualise the overall market trajectory.

While this analysis utilises short-term rental data (Airbnb) as a proxy for housing market value, previous studies suggest a strong positive correlation between short-term nightly rates and long-term rental yields within metropolitan areas. Therefore, the spatial trends identified herein are considered indicative of the broader residential market.

Appendix B: Source Code

```
# ==============================================================================
# PROJECT: THE BRISBANE "COMMUTER TAX" ANALYSIS
# AUTHOR: April Williams
# DATE: December 2025
# ==============================================================================

# ~ 1. SETUP & IMPORTS ~
import pandas as pd
import numpy as np
import matplotlib.pyplot as plt
import seaborn as sns

# Set professional chart style
sns.set_theme(style="whitegrid")
```

```
# ~ 2. DATA LOADING & CLEANING ~
# Load the raw data
try:
    df = pd.read_csv('listings.csv')
    print("Y File loaded successfully.")
except FileNotFoundError:
    print("N Error: 'listings.csv' not found. Please upload it first.")

# Clean the Price Column (Remove '$' and ',')
# Create a copy to avoid warnings
df_clean = df.copy()

# Remove symbols and convert to float (Force it to be a number)
df_clean['price'] = df_clean['price'].astype(str).str.replace(r'[$,]', '',
regex=True)
df_clean['price'] = pd.to_numeric(df_clean['price'], errors='coerce')

# Drop missing values in critical columns
df_clean = df_clean.dropna(subset=['latitude', 'longitude', 'price'])

# Filter outliers (Remove listings > $1,000 to see the "real" market)
df_clean = df_clean[df_clean['price'] < 1000]

print(f"Y Data cleaned. Analysis sample size: {len(df_clean)} listings.")

# ~ 3. FEATURE ENGINEERING (DISTANCE CALCULATION) ~
# Define Brisbane CBD Coordinates
cbd_lat = -27.4698
cbd_lon = 153.0251

# Define Haversine Function (Calculate distance on round Earth)
def haversine_distance(lat1, lon1, lat2, lon2):
    R = 6371  # Earth radius in km
    phi1, phi2 = np.radians(lat1), np.radians(lat2)
    dphi = np.radians(lat2 - lat1)
    dlambda = np.radians(lon2 - lon1)
    a = np.sin(dphi/2)**2 + np.cos(phi1)*np.cos(phi2) * np.sin(dlambda/2)**2
    return 2 * R * np.arctan2(np.sqrt(a), np.sqrt(1 - a))

# Apply to dataset
df_clean['dist_cbd_km'] = haversine_distance(
    df_clean['latitude'],
    df_clean['longitude'],
    cbd_lat,
    cbd_lon
)

# ~ 4. VISUALISATION 1: THE GEOSPATIAL MAP ~
plt.figure(figsize=(10, 8))

# Create Scatter Plot
sns.scatterplot(
    data=df_clean,
    x='longitude',
    y='latitude',
    hue='price',
    size='price',
    sizes=(5, 100),       # Dots range from size 5 to 100
    palette='viridis_r',  # Reversed: Yellow = Expensive
```

```
    alpha=0.7,
    legend=False          # We build a custom colorbar below
)

# Create Professional Color Bar
norm = plt.Normalize(df_clean['price'].min(), df_clean['price'].max())
sm = plt.cm.ScalarMappable(cmap="viridis_r", norm=norm)
sm.set_array([])
cbar = plt.colorbar(sm, ax=plt.gca(), label='Price per Night ($)')

# Labels & formatting
plt.title('Brisbane Rental Landscape (2025)', fontsize=14, weight='bold')
plt.xlabel('Longitude')
plt.ylabel('Latitude')
plt.axis('equal') # Prevent map stretching
sns.despine(left=True, bottom=True)

plt.show() # Display Map

# ~ 5. VISUALISATION 2: THE "COMMUTER TAX" PROOF ~
plt.figure(figsize=(10, 6))

# Scatter Plot with Regression Line
sns.regplot(
    data=df_clean,
    x='dist_cbd_km',
    y='price',
    scatter_kws={'alpha': 0.3, 'color': 'purple', 's': 10}, # Light purple dots
    line_kws={'color': 'red', 'linewidth': 2}               # Strong red trend line
)

# Labels
plt.title('The Commuter Tax: Price vs. Distance to CBD', fontsize=14,
weight='bold')
plt.xlabel('Distance from CBD (km)')
plt.ylabel('Price per Night ($)')
plt.xlim(0, 50) # Focus on the inner 50km

plt.show() # Display Graph
```