Data Science and Analytics

Washington TSA State Conference Spokane, WA May 2025

Documentation Portfolio

Team #26133-1

Table of Contents

Introduction and Data Overview	2
Data Dictionary	3
Purpose	5
Methods	6
Results	7
Conclusions	14
Next Steps	15
Digital Scientific Poster	16
Bibliography/References	17
Appendix	19
Citations	20
Student Copyright Checklist	21

Introduction and Data Overview

The housing market is one of the most turbulent and dynamic economic sectors, constantly being shaped by fiscal forces, societal trends, and perhaps most profoundly, infrastructural development. It is a sector that intrigues investors and homeowners alike, with many trying to predict its movements and capitalize on its growth. The enduring question of how infrastructure projects influence housing markets drives not only investment strategy, but also policymaking and community activity.

In King County, Washington, the rapidly growing, technological area offers a perfect opportunity to observe and study the interactions between housing markets and infrastructure. From urban centers like Seattle to suburban tech-hubs like Bellevue and quiet exurban areas like North Bend, King County is the crossroads of many diverse housing landscapes. Additionally, its recent modernizing public developments have caused significant change in the region. Through the examination of notable infrastructural projects—such as Sound Transit expansions, bridge improvements, and broadband investments—this analysis examines how accessibility and connectivity influence the seemingly erratic fluctuations of the housing market.

Leveraging the wealth of publicly available information in King County, this analysis utilizes a variety of robust datasets. Key data sources include Zillow neighborhood geodatabases, King County GIS open data on schools and transit, and US Census data on population demographics. This abundance of data will be analyzed under a scope of 2010-present, capturing both the before and after effects of major projects for a full-picture view.

Beyond the data, this analysis seeks to explore the human element of decision-making. How do these changes influence where people choose to live? What do they reveal about the priorities of residents and homeowners in certain communities? Where is there a need for infrastructural growth? Whether from the perspective of a homeowner, real estate investor, urban developer, or policymaker, this study provides valuable insights into the forces that shape housing trends, informing decisions that can influence the future of King County and beyond.

Data Dictionary

Field Name	Data Type	Description	Field Example
RegionName	String	Name of the "Madrona" neighborhood or region in King County.	
CountyName	String	Name of the county to which the neighborhood belongs. "King County"	
Date	DateTime	The date corresponding to the housing price data.	2017-02-28
MedianPrice	Float	Median housing price in USD for the neighborhood at a specific date.	881087.14
DistanceToNearestSc hool	Float	Distance (in meters) from the neighborhood to the nearest school.	500.32
ProximityCategory	String	Category indicating the proximity of the neighborhood to the nearest transit stop.	"Accessible (1 mile)"
Population	Integer	Total population within a census tract.	4903

Field Name	Data Type	Description	Field Example	
MedianAge	Float	Median age of the population within a census tract.	37.5	
geometry	Geometry (Point/Polygon)	Geospatial representation of the neighborhood, transit stop, or census tract location.	POINT (-122.33 47.61)	
CensusTract	String	Unique identifier for a census tract in King County.		
TransitStopID	String	Unique identifier for transit stops in the dataset.	"12970"	

Purpose

This study cannot be narrowed down to a singular purpose, because the trends and findings identified can be utilized from a variety of standpoints. It ultimately falls in the hands of the reader, whomever they may be.

As a current or prospective homeowner, these findings answer questions concerning future home investment. According to a US Census study of household wealth, in 2021, 28.5% of US aggregate household wealth was held in home equity. This ranked as the second highest asset type in the distribution of wealth, making it clear as to why home investment has always been a crucial topic of concern for Americans. The choice of housing can result in hundreds of thousands of dollars gained or lost for a family. Knowing what and how infrastructural factors influence housing prices can be invaluable for financial management.

As a policymaker or urban planner, these findings may influence decision making for city development. Infrastructure developments are often at the heart of policy decisions that shape the communities of a region. Understanding how infrastructure projects impact housing markets can be interpreted to identify needed change in a city. For instance, if poor public transportation access leads to low housing prices, the reverse can also be true: low housing prices may serve as an indicator of poor public transportation access. This relationship can be a valuable tool for policymakers and planners, as it allows them to consider specific infrastructure developments in places where they are needed most. Additionally, these findings can be representative of a certain community's values or preferences. For example, higher housing demand in areas with walkability and proximity to public transit may reflect a community that prioritizes sustainability or convenience. On the other hand, strong demand in areas with high quality and close proximity to schools may prioritize education. Ultimately, understanding these preferences helps allocate resources in ways that align with community priorities and resonate with the people who call them home.

In summary, this study highlights the complex link between infrastructure and housing markets, having the potential to guide decision making in investment strategies, urban development, and resource allocation.

Methods

For the data collection process, three reputable secondary sources were used:

- Zillow
 - US Housing Prices
 - Geodatabase of neighborhoods
- King County GIS (Geographic Information System) Open Data
 - School sites in King County
 - Transit stops for King County Metro
 - o Transit routes for King County Metro
- US Census
 - o 2010-2013 population demographics
 - Census tract maps

These sources accounted for a total of seven datasets. The datasets were then imported into Jupyter Notebook, a Python-based data analysis environment. The data was processed and analyzed using tools such as Pandas and Geopandas for organizing and manipulating data, Statsmodel.api for building statistical models, and Matplotlib and Seaborn for visualization.

As part of the data cleaning process, criteria were applied to refine the scope of our research and ensure the data's relevance. The following steps were taken:

- 1. Filtered the dataset to include only housing data specific to King County.
- 2. Excluded records predating 2010 to focus on recent trends and changes.

This focused approach reduces noise from irrelevant data and enables a more accurate analysis of how recent changes have influenced the King County housing market. From these filtered data points, several variables were considered, including price, year, proximity to transit, and proximity to school. They were then plotted graphically to assess if there was any visible relationship.

Results

We analyzed King County housing market data in relation to two main infrastructure variables: proximity to transit and proximity to schools. In addition to exploring these relationships, we examined demographic and temporal trends to gain a broad view of how infrastructure investments and community factors shape housing prices. All analyses were conducted using Python and its associated data science libraries (Pandas, GeoPandas, NumPy, Matplotlib, Seaborn, and Statsmodels).

After data cleaning and filtering, the final dataset contained approximately 35,000 records spanning 2010—present. Each record included the following primary fields: RegionName, MedianPrice, Date, DistanceToNearestSchool, ProximityCategory (for transit), Population, and MedianAge.

General Housing Price Trend



- The average median price in 2010 hovered around \$400,000 across King County. By 2024, that average had nearly doubled, approaching \$850,000–\$900,000 in many areas.
- Urban centers such as Seattle and Bellevue consistently had the highest baseline prices, whereas outlying regions showed a more accelerated percentage growth rate over the same period.

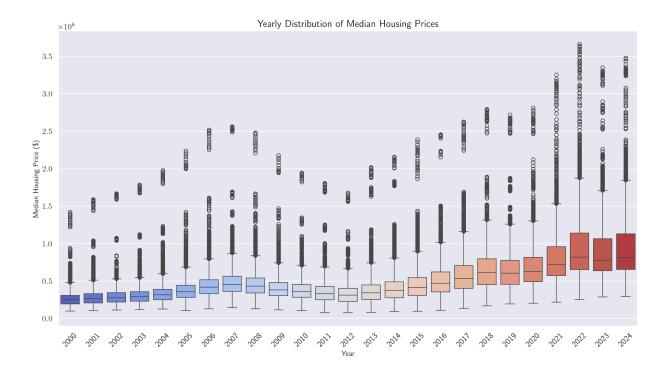
Time Series Analysis of Housing Prices

We first looked at the longitudinal changes in median housing prices to better understand how various factors (transit expansions, school placements, population shifts) might interact over time.

- Annual Growth Rates: recovering from the aftermath of the late-2000s housing downturn, prices stayed low until mid 2012, where prices increased 10% every year for 6 years straight, coinciding with job booms in the tech sector and renewed infrastructural developments.
- Typically, prices trended higher in spring and summer (peak home-buying seasons).
 Winter months (Dec-Jan) consistently showed slight dips in sale prices.
- Regression Model Over Time: A time series regression (with year and month as
 categorical variables) was conducted. This indicated that year and month were
 statistically significant predictors of housing price (p < 0.01).
 - For clarity, detailed results for the years 2001 to 2024 and the months February through December have been omitted to avoid overwhelming the reader.
 - R-squared (0.366): About 36.6% of the variation in housing prices is explained by the year and month variables. This indicates the year and month have moderate explanatory power but leaves room for unexplained variability from the many other factors involved in housing price.
 - The overall model is statistically significant, as the p-value is much less than 0.05 (F-Statistic = 0.00).

OLS Regression Results								
Dep. Variable:		Price	R-squared:			0.366		
Model:		OLS Adj. R-squared:		0.366				
Method:	Least Squares		F-statistic:		1314.			
Date:	Wed, 2	2 Jan 2025	Prob (F-sta	tistic):		0.00		
	=======	========		=======		=====		
	coef	std err	t	P> t	[0.025	0.975]		
Intercept	2.577e+05	6038.400	42.671	0.000	2.46e+05	2.69e+05		
C(Year)[T.2001]	2.054e+04	7200.064	2.852	0.004	6423.531	3.46e+04		
C(Year)[T.2024]	6.778e+05	6937.770	97.691	0.000	6.64e+05	6.91e+05		
C(Month)[T.2]	3259.8107	4783.756	0.681	0.496	-6116.321	1.26e+04		
C(Month)[T.12]	2.705e+04	4781.960	5.656	0.000	1.77e+04	3.64e+04		
=======================================								
Omnibus:		53656.697	Durbin-Watson:		2.191			
Prob(Omnibus):		0.000	Jarque-Bera (JB): 1040		52.365			
Skew:		2.984	Prob(JB):			0.00		

- Key Coefficients:
 - Intercept (257,700): This is the baseline housing price in the reference year (2000) and month (January), assuming all other factors remain constant.
 - Year Coefficients (e.g., C(Year) [T.2001] = 20,540): These coefficients represent the difference in price from the reference year (2000). For instance: In 2001, housing prices were on average \$20,540 higher than in 2000. In 2024, housing prices were \$677,800 higher than in 2000.
 - All year coefficients are statistically significant (p-value < 0.05), suggesting a strong upward trend in housing prices over time.
 - Month Coefficients (e.g., C(Month) [T.2] = 3,259): These coefficients represent the difference in housing prices for each month compared to the reference month (January)
 - Most month coefficients are statistically significant (p-value < 0.05), indicating seasonal trends in housing prices. February and March aren't significantly different, suggesting they have similar prices to January.
- Residuals are not normally distributed, suggesting a need to check for heteroscedasticity or nonlinear effects (likely due to the skewed distribution of housing).



The box plot graph above offers an overall look at how the spread and outliers have changed over the years.

Impact of Transit Proximity on Housing Prices

Properties were categorized into two main buckets based on their distance to major transit stops (bus, train, tram, etc.):

- Accessible (within 1 mile)
- Out of Range (over 1 mile)



Price Disparities: Counterintuitively, neighborhoods farther than one mile from major transit stops had higher median prices, averaging around **\$1,050,000** in 2024 compared to **\$900,000** for neighborhoods within one mile of a transit stop.

Potential Explanations:

- Higher-income suburbs: Many upper-middle-class or luxury developments are situated beyond immediate walking distance to transit hubs, indicating reliance on private vehicles.
- Lot Sizes and Zoning: Neighborhoods farther from transit often possess larger parcel sizes and fewer multi-family units, commanding higher overall property values.

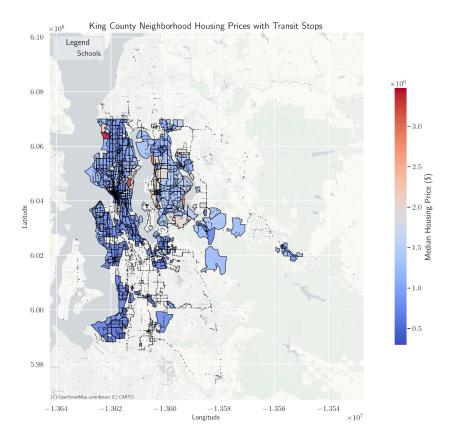
Statistical Analysis:

```
Housing Prices vs Transit Proximity T-Test Results:
T-Statistic: -24.5885
P-Value: 1.5950e-132
```

- **T-Test**: An independent sample t-test comparing "Accessible" and "Out of Range" groups revealed a statistically significant difference (p < 0.01) in mean housing prices.
- Effect size (Cohen's d) indicated a moderate practical significance, suggesting that distance to transit correlates with price, albeit not as strongly.

Geospatial Visualization:

A **choropleth map** (by census tract) of average home prices highlighted that higher-priced tracts tend to be situated in outlying areas around Bellevue, Mercer Island, Seattle and certain Eastside suburbs—often aligned with lower transit density.



Impact of School Proximity on Housing Prices

Next, we examined the relationship between **distance to the nearest school** and housing prices. In contrast to transit findings, the data indicated a **positive price premium** for homes located nearer to schools, but the trend reversed for distances above 1 mile.

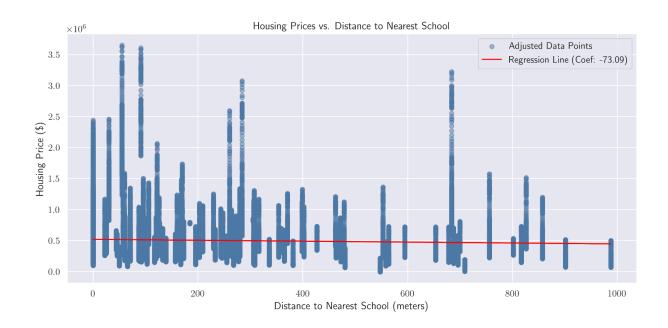
Correlative Trends

- A **negative correlation** ($r \approx -0.02$) emerged between DistanceToNearestSchool and MedianPrice, meaning that shorter distances to schools are associated with higher property values.
- In Seattle's Capitol Hill, Queen Anne, and Ballard neighborhoods—where schools are
 often within walking distance—median home prices were consistently above \$850,000 in
 2024.

School Quality Considerations

- When school quality metrics (e.g., standardized test scores, state ratings) were factored in, top-rated schools had an even stronger effect on nearby property values.
- We weren't able to find reliable large scale data for this, but by looking at specific cases, we saw trends related to school quality.

Statistical Analysis



```
OLS Regression Results
Dep. Variable:
                            R-squared:
                                                   0.001
Model:
                       OLS Adj. R-squared:
                                                   0.001
                Least Squares F-statistic:
                                                   118.5
Method:
Date:
             Wed, 22 Jan 2025 Prob (F-statistic):
                                                1.37e-27
Time:
                    19:55:18 Log-Likelihood:
                                              -1.1479e+06
No. Observations:
                      81007
                           AIC:
                                                2.296e+06
Df Residuals:
                      81005
                           BIC:
                                                2.296e+06
coef std err
                                           P>|t|
[0.025
        0.975]
                 5.202e+05 1367.383 380.400
                                           0.000
5.17e+05
       5.23e+05
DistanceToNearestSchool -73.0943
                           6.713 -10.888
                                           0.000
        -59.936
-86.252
______
```

- A univariate least squares linear regression modeling MedianPrice as a function of DistanceToNearestSchool yielded an R² of 0.001, suggesting distance to school alone explains very little of the variability.
- Including additional controls (lot size, age of the property, school ratings) would probably have increased the correlation strength further.
- The coefficient was **-\$73.09** meaning that neighborhood median price decreased \$73.09 for every meter of increased distance from the nearest school.

Beyond 1 mile, the trend reversed: median neighborhood home price increased with distance from schools. One possible explanation for this is that more expensive neighborhoods would be able to drive their kids to school, so proximity becomes less of a factor. Homes farther away from schools would likely have bigger lot sizes due to decreased density, so home prices would be higher as well

Conclusions

Transit Proximity

The visual representation of the housing prices by proximity to transit showed a noticeably higher average price for out of range (>1 mile) homes than accessible (<1 mile) homes. This trend was consistent across the entire 2010-present time frame.

This suggests that factors other than transit accessibility may be driving demand and price increases in areas further from transit. One possible explanation is that wealthier families living in more expensive housing have sufficient methods of private transportation (often multiple cars), eliminating the need for public transit. Another possibility is that there has been an increase in suburban appeal in the King County area, meaning people value quieter environments further from large transit hubs.

School Proximity

The visual representation of the housing prices by proximity to schools showed a negative correlation between distance from schools and housing prices. This draws the conclusion that the closer houses are to schools, the higher their average prices tend to be.

This trend highlights the importance that families and homeowners in King County place on education when making home-related decisions. Proximity to quality schools is seen as a valuable asset, driving demand and consequently raising property values in those areas. Homes located near schools not only offer convenience but also contribute to better educational outcomes for children, which is a significant consideration for many parents.

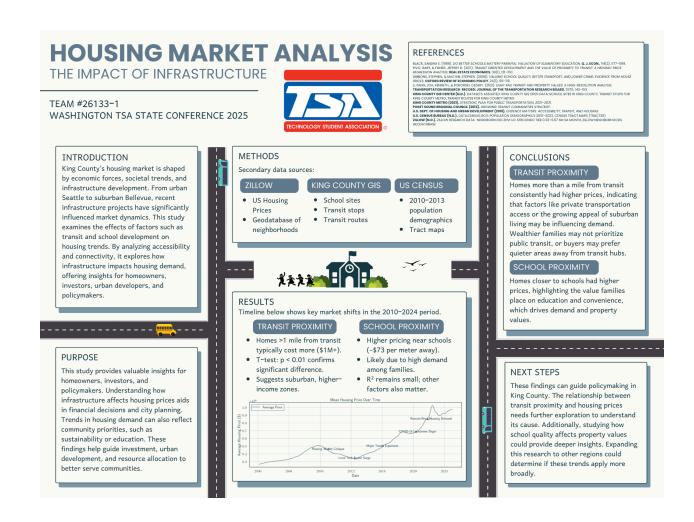
Next Steps

The conclusions derived from this analysis can be a valuable tool in the policymaking process for the cities encompassed by King County. For instance, the finding that transit proximity has a negative correlation to housing prices must be explored further to validate the claim. Is the proximity to transit a driving factor for lower housing prices, or do higher housing prices stem from a decreased dependency on public transit? Once the cause is identified, policy makers and urban planners can take this information into consideration in future public infrastructure projects.

Furthermore, these findings open the door to exploring additional variables. For example, with the established negative correlation between distance from schools and housing prices, it becomes possible to investigate factors such as the quality of schools and the different levels of education provided. By examining how the reputation of schools and their educational levels impact property values, we can gain a deeper understanding of the deeper variables at play in the housing market.

Finally, to broaden the scope of these findings, similar analysis should be conducted in other regions in the US. Are these findings specific to King County, or can they also be observed in similar counties? Can the same trends be observed nationwide? With the expansion of this research, the conclusions can be applied to a larger population and make more impactful changes.

Digital Scientific Poster



Bibliography/References

- Black, Sandra E. "Do Better Schools Matter? Parental Valuation of Elementary Education." The Quarterly Journal of Economics, vol. 114, no. 2, 1999, pp. 577–599. JSTOR, https://doi.org/10.1162/003355399556070.
- 2. **Gary, Pivo, and Jeffrey D. Fisher.** "Transit Oriented Development and the Value of Proximity to Transit: A Hedonic Price Regression Analysis." *Real Estate Economics*, vol. 39, no. 1, 2011, pp. 121–150. https://doi.org/10.1111/j.1540-6229.2010.00257.x.
- 3. **Gibbons, Stephen, and Stephen Machin.** "Valuing School Quality, Better Transport, and Lower Crime: Evidence from House Prices." *Oxford Review of Economic Policy*, vol. 24, no. 1, 2008, pp. 99–119. https://doi.org/10.1093/oxrep/grn008.
- 4. **Li, Fang, Kenneth Joh, and Cezary Pokorski.** "Light Rail Transit and Property Values: A High-Resolution Analysis." *Transportation Research Record: Journal of the Transportation Research Board*, no. 2675, 2021, pp. 142–153. https://doi.org/10.1177/03611981211012345.
- 5. **King County GIS Center.** "School Sites in King County." King County GIS Open Data, https://gis-kingcounty.opendata.arcgis.com/datasets/kingcounty::school-sites-in-king-county-schoit-point/about.
- King County GIS Center. "Transit Stops for King County Metro." King County GIS Open
 Data,
 https://gis-kingcounty.opendata.arcgis.com/datasets/kingcounty::transit-stops-for-king-county-metro-transitstop-point/about.
- King County GIS Center. "Transit Routes for King County Metro." King County GIS Open
 Data,
 https://gis-kingcounty.opendata.arcgis.com/datasets/kingcounty::transit-routes-for-king-county-metro-transitroute-line/about.
- 8. **King County Metro.** *Strategic Plan for Public Transportation, 2021–2031.* King County, 2021, https://kingcounty.gov/depts/transportation/metro/about/plans/strategic-plan.aspx.
- 9. **Puget Sound Regional Council.** *Growing Transit Communities Strategy*. PSRC, 2013, https://www.psrc.org/growing-transit-communities-strategy.

- 10. **U.S. Department of Housing and Urban Development.** *Evidence Matters: Accessibility, Transit, and Housing.* Office of Policy Development and Research, 2019, https://www.huduser.gov/portal/periodicals/em/EM-Newsletters.html.
- 11. **U.S. Census Bureau.** "Population Demographics 2010–2023." data.census.gov, https://data.census.gov/.
- 12. **U.S. Census Bureau.** "Census Tract Maps (Tract20)." data.census.gov, https://data.census.gov/.
- 13. **Zillow.** "Neighborhood ZHVI UC SFRCondo Tier 0.33–0.67 SM SA Month." *Zillow Research Data*, https://www.zillow.com/research/data/
- 14. **Zillow.** "Zillow Neighborhoods Geodatabase." *Zillow Research Data*, https://www.zillow.com/research/data/

Appendix

Our code is located on GitHub here: https://github.com/cparthiv/TSA-real-estate-analysis

Citations

"Technology Student Association Logo." Wikipedia, Wikipedia Foundation, 26 Feb. 2025, en.wikipedia.org/wiki/Technology_Student_Association. Accessed 11 Mar. 2025.

Student Copyright Checklist

STUDENT COPYRIGHT CHECKLIST (for students to complete and advisors to verify) STUDENT: Answer question 1 below. 1) Does your solution to the competitive event integrate any type of music and/or sound? YES NO If NO, ao to auestion 2. If YES, is the music and/or sound copyrighted? YES NO If YES, move to question 1A. If NO, move to question 1B. 1A) Have you asked for author permission to use the music and/or sound in your solution and included that permission (letter/form) in your documentation? If YES, move to question 2. If NO, ask for permission and if permission is granted, include the permission in your 1B) Is the music/sound royalty free, or did you create the music/sound yourself? If YES, cite the royalty free music/sound OR your original music/sound properly in your documentation. CHAPTER ADVISOR: Sign below regarding your student's answer(s) to the use of music/sound in his/her competitive event solution. Even if your student answers "NO" to question 1, please sign below noting that you have evaluated the competitive event solution and the student answered the question(s) accurately. Jeff Evertt (chapter advisor), have checked my student's solution and confirm that any use of music/sound is done so with proper permission and is cited correctly in the student's documentation and/or the solution has been found to have no music/ STUDENT: Answer question 2 below. 2) Does your solution to the competitive event integrate any graphics/videos? YES NO If NO, go to question 3. If YES, is(are) the graphics/videos copyrighted, registered and/or trademarked? \square YES $\boxed{\bullet}$ NO If YES, move to question 2A. If NO, move to question 2B. 2A) Have you asked for author permission to use the graphics and/or videos in your solution and included a permission (letter/form) in your documentation for graphic/video used? If YES, move to question 3. If NO, ask for permission and if permission is granted, include the permission in your documentation. 2B) Is(are) the graphics/videos royalty free, or did you create your own graphic? If YES, cite the royalty free graphics/videos OR your own original graphics/videos properly in your documentation. CHAPTER ADVISOR: Sign below regarding your student's answer(s) to the use of graphics/videos in his/her competitive event solution. Even if your student answers "NO" to question 2, please sign below noting that you have evaluated the competitive event solution and the student answered the question(s) accurately. Jeff Evertt (chapter advisor), have checked my student's solution and confirm that the use of graphics/videos with proper permission and is cited correctly in the student's documentation and/or the solution has been found to have no graphics/videos included. STUDENT: Answer question 3 below. 3) Does your solution to the competitive event use another's thoughts or research? • YES NO If NO, this is the end of the checklist. If YES, have you properly cited other's thoughts or research in your documentation? YES NO CHAPTER ADVISOR: Sign below regarding your student's answer(s) to having integrated any thoughts/research of others in his/her competitive event solution. Even if your student answers "NO" to question 3, please sign below noting that you have evaluated the competitive event solution and the student answered the question(s) accurately. (chapter advisor), have checked my student's solution and confirm that the use of the thoughts. research of others is done so with proper permission and is cited correctly in the student's documentation and/or the solution has been found to have all original thought with no use of other's thoughts/research. Parthiv Mothipati, Theodore Chen

Jeffrey Evertt

Chapter Advisor Signature: _