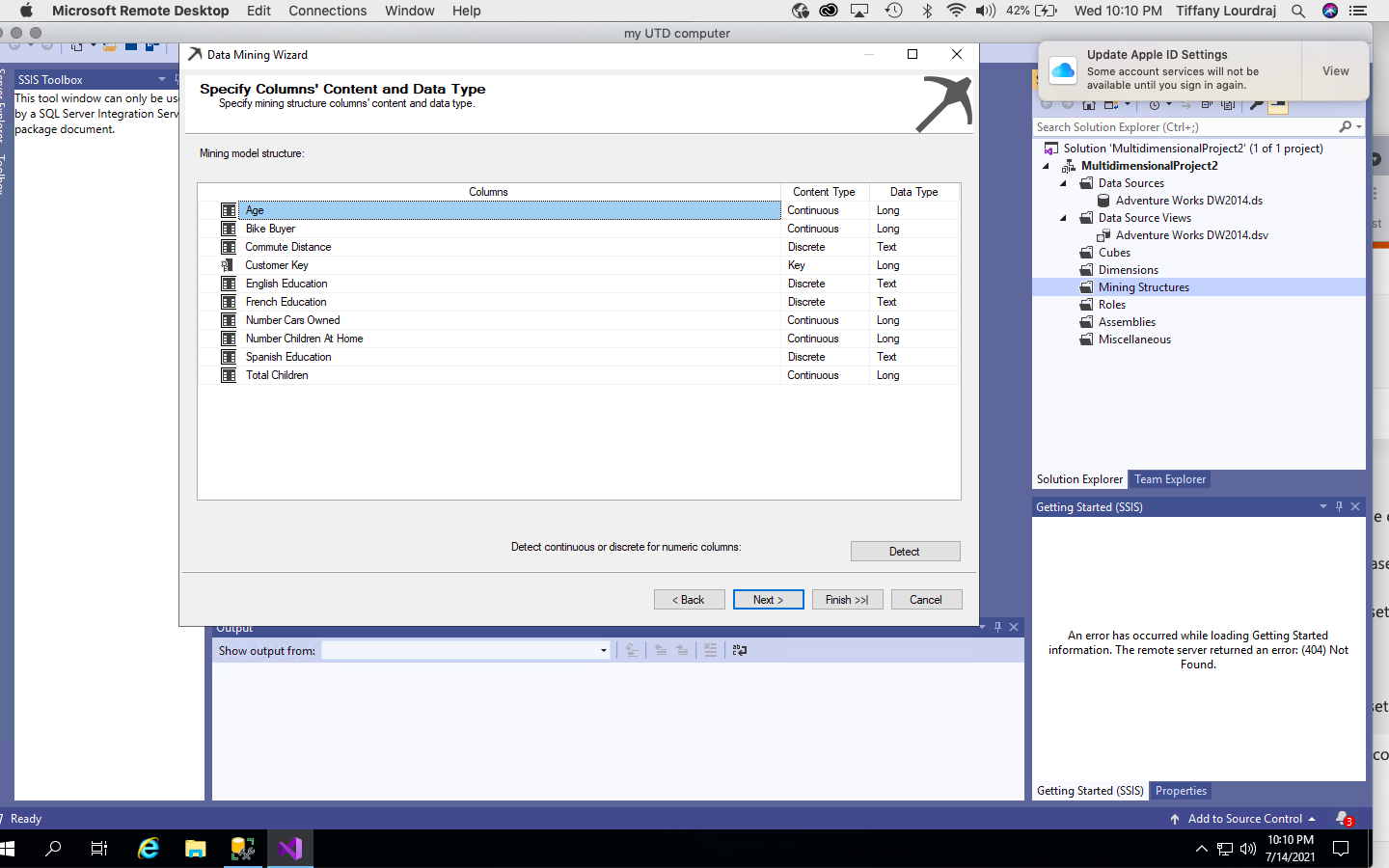
**ITSS 4351 Foundations of Business Intelligence**

**Mining Models**

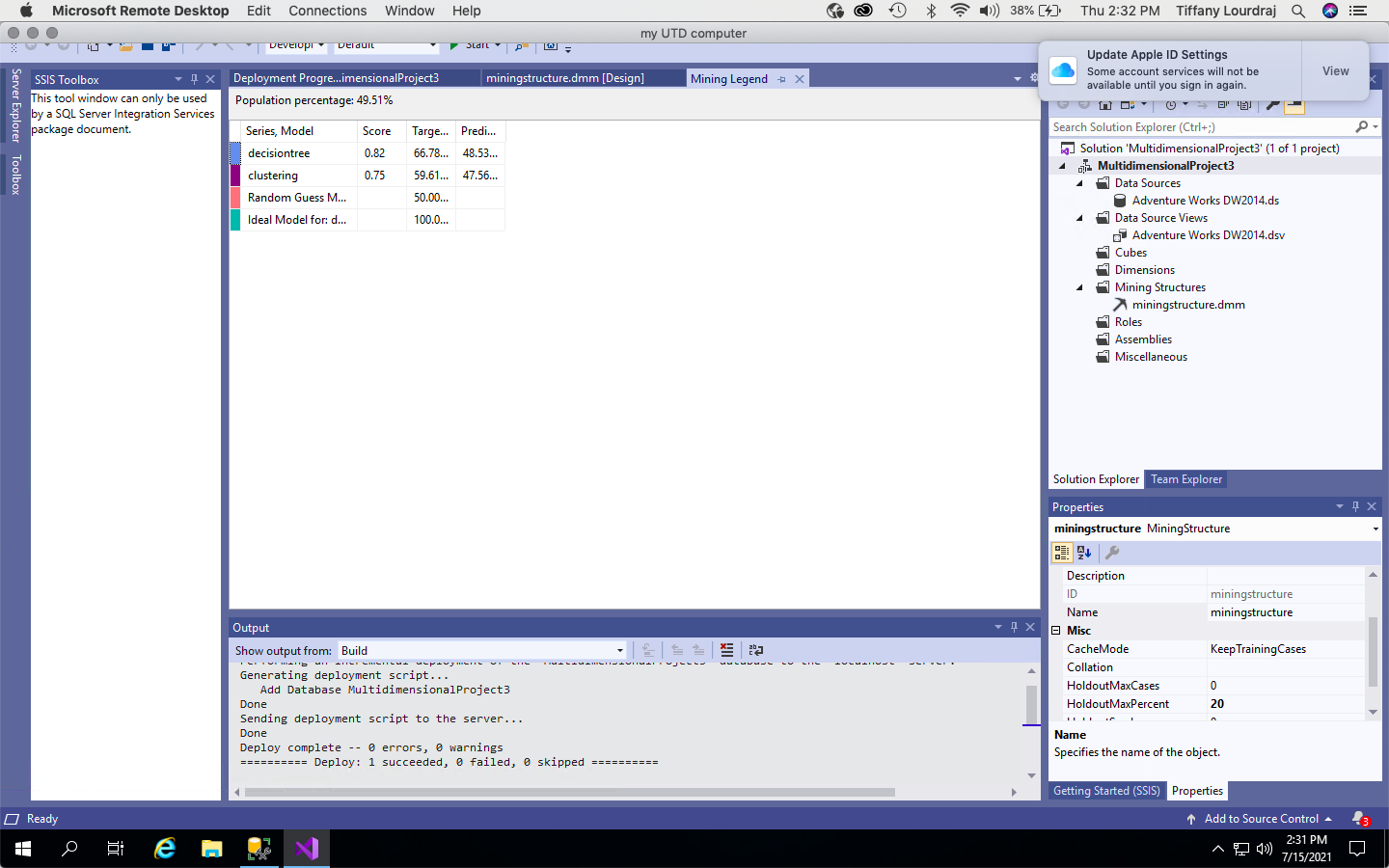
Note:

* 1. Use the AdventureWorks2014 Database and the vTargetMail view.
  2. Partial Screenshots will not be evaluated. Your screenshots should cover the entire screen and prominently show your work
  3. **ONLY ONE SUBMISSION** is allowed. Please ensure that you submit the assignment carefully.
  4. No Late Submission is allowed. The key will be made available right after the submission deadline and hence no late submissions will be allowed.

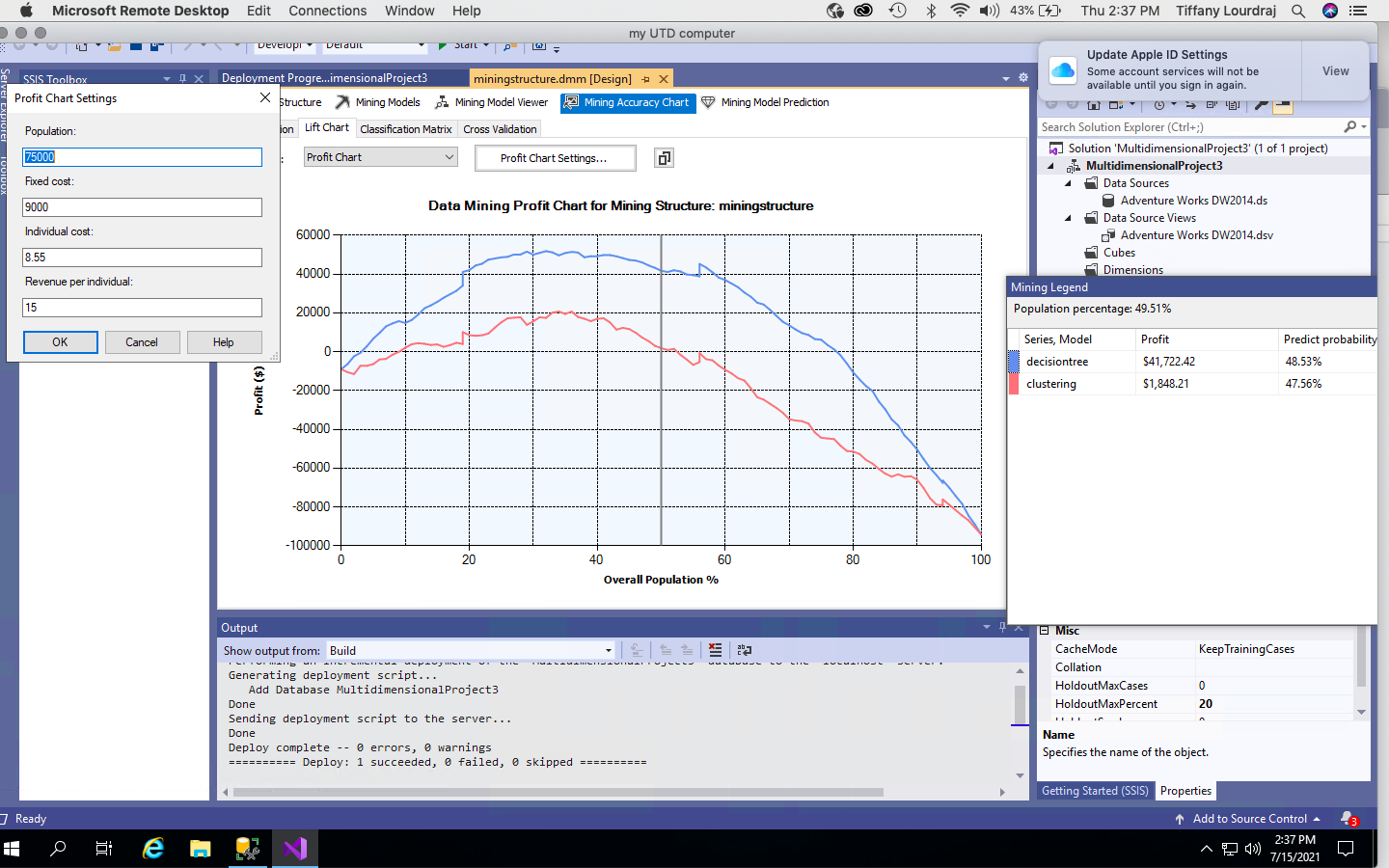
1. You are required to predict the BikeBuyer quotient using any input which has a Score of at least 0.02. **Paste a Screenshot of all the selected inputs - 10 points**
2. Ensure the Content Type is aligned with what Visual Studio recommends. Using 20% data for testing and Allowing Drill Through, create a Decision Tree. Compare it with the Clustering Model and find out which one has a better score, for BikeBuyer = 1. What are the scores? **Paste a screenshot of the mining legend – 10 points**

**decisión tree has a higher score and is a better prediciton for the bike buyers**

**82% decision tree**

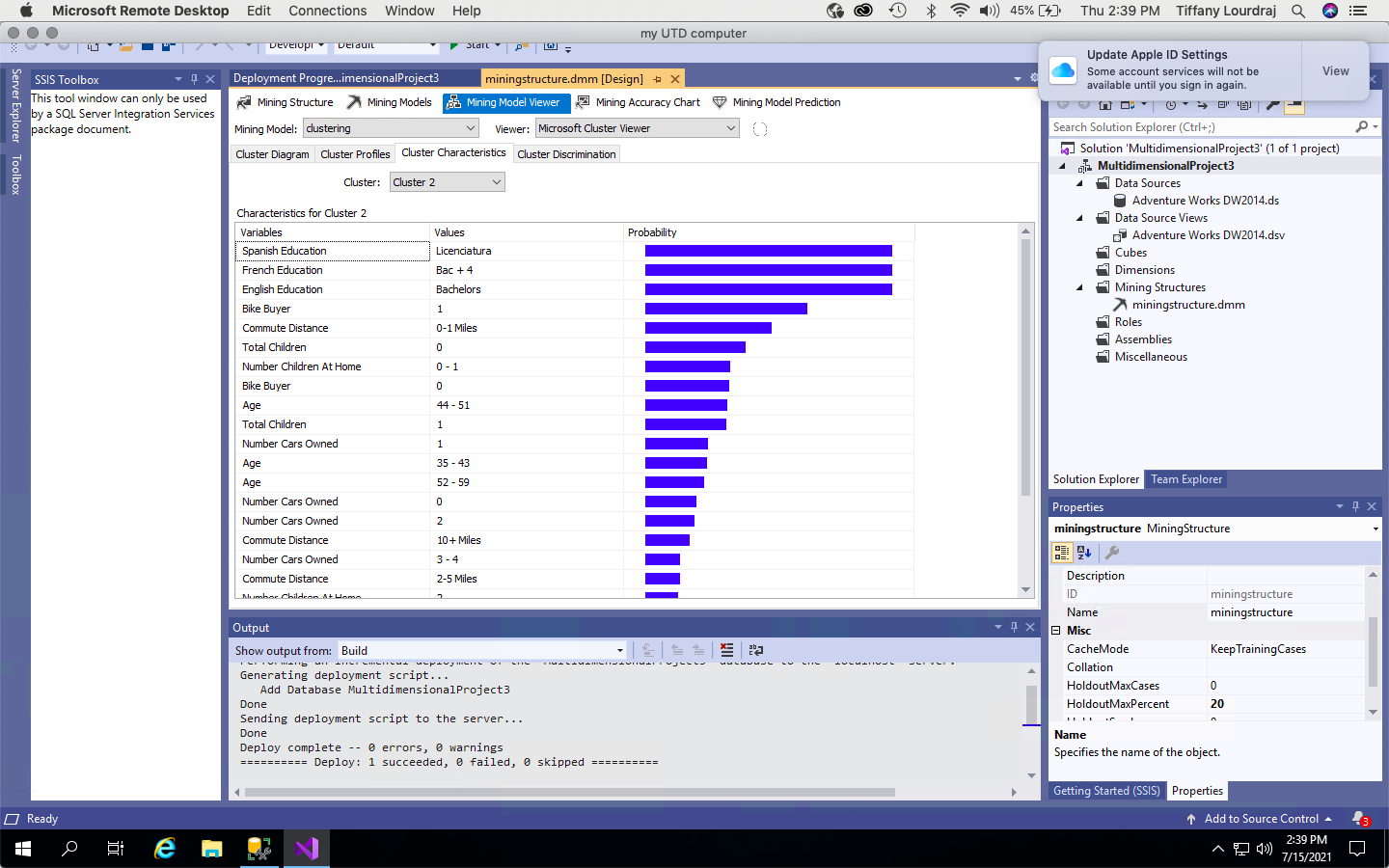
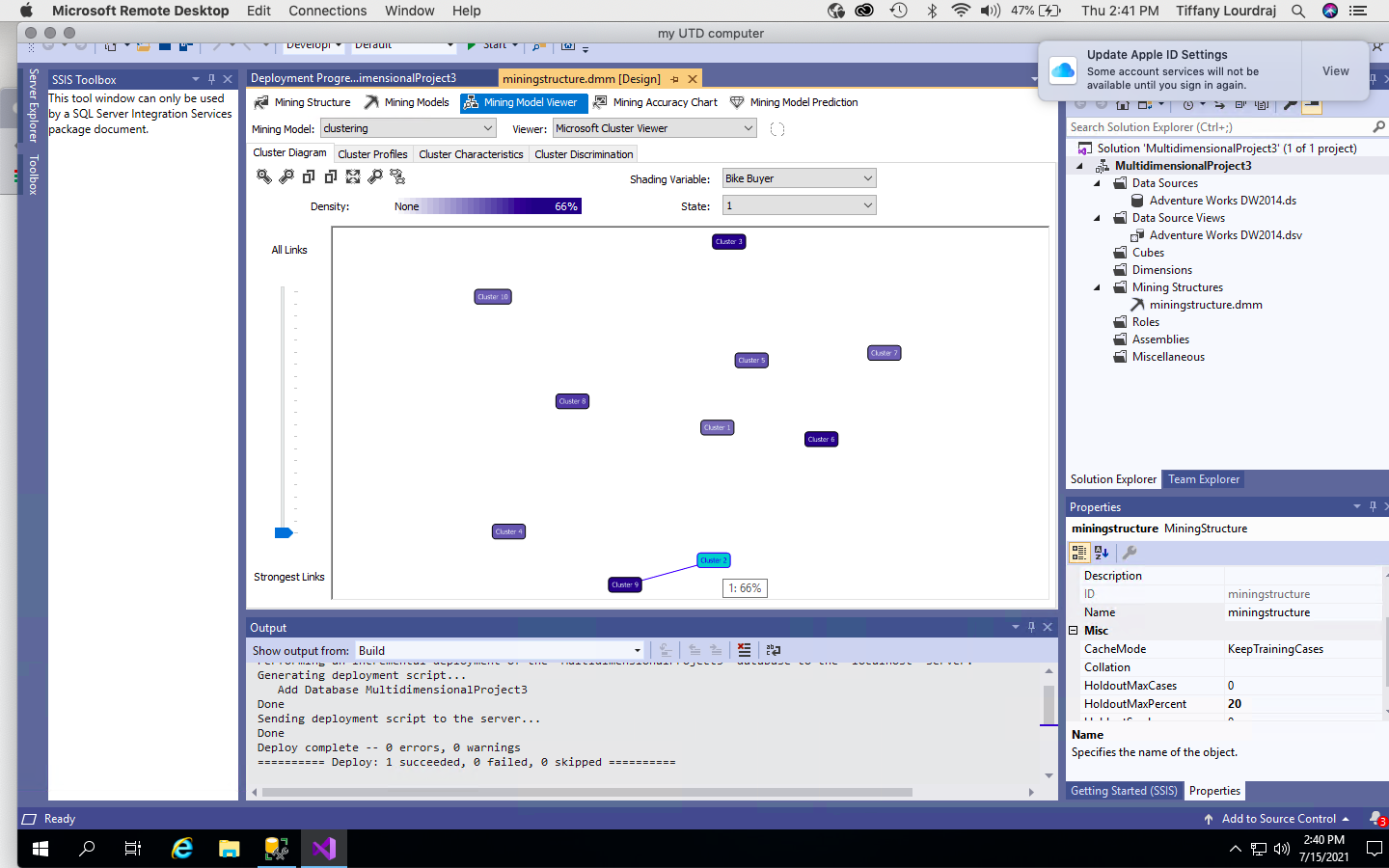
**75% clustering **

1. If I live in a Town with the population of 75,000, using the Decision Tree, if my Fixed Cost is $9000, and I want to earn $15 per individual, what is the maximum I can spend on an individual so that I don’t make a loss. **Paste the screenshot of the ‘Profit Chart Settings’ and the Mining Legend – 20 points**

****

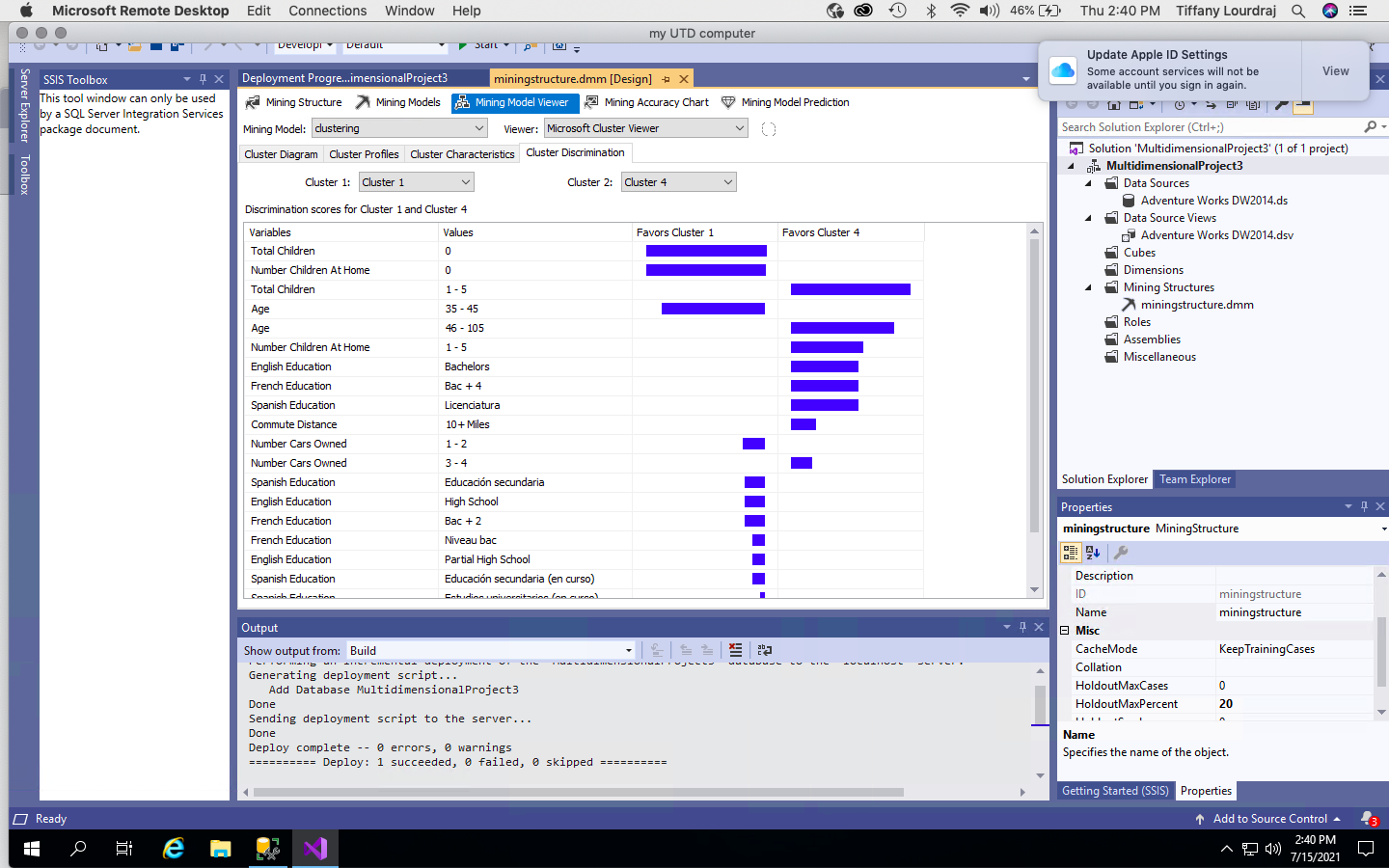
1. The Cluster which has the maximum percentage of people who bought a bike. Describe this cluster using 4 inputs. **Paste a screenshot of this cluster and** **describe your findings in at least 30 words. – 20 points**

Cluster 2 has the the greatest density which is the highest amount which is equivalent to 66%. you will find the highest density when you find the darkest colored circle. This is why I chose cluster 2 for the maximum percentage of people who bought a bike. Spanish education, french education and english education and commute distance all have a major impact on the biker buyers decision.

****

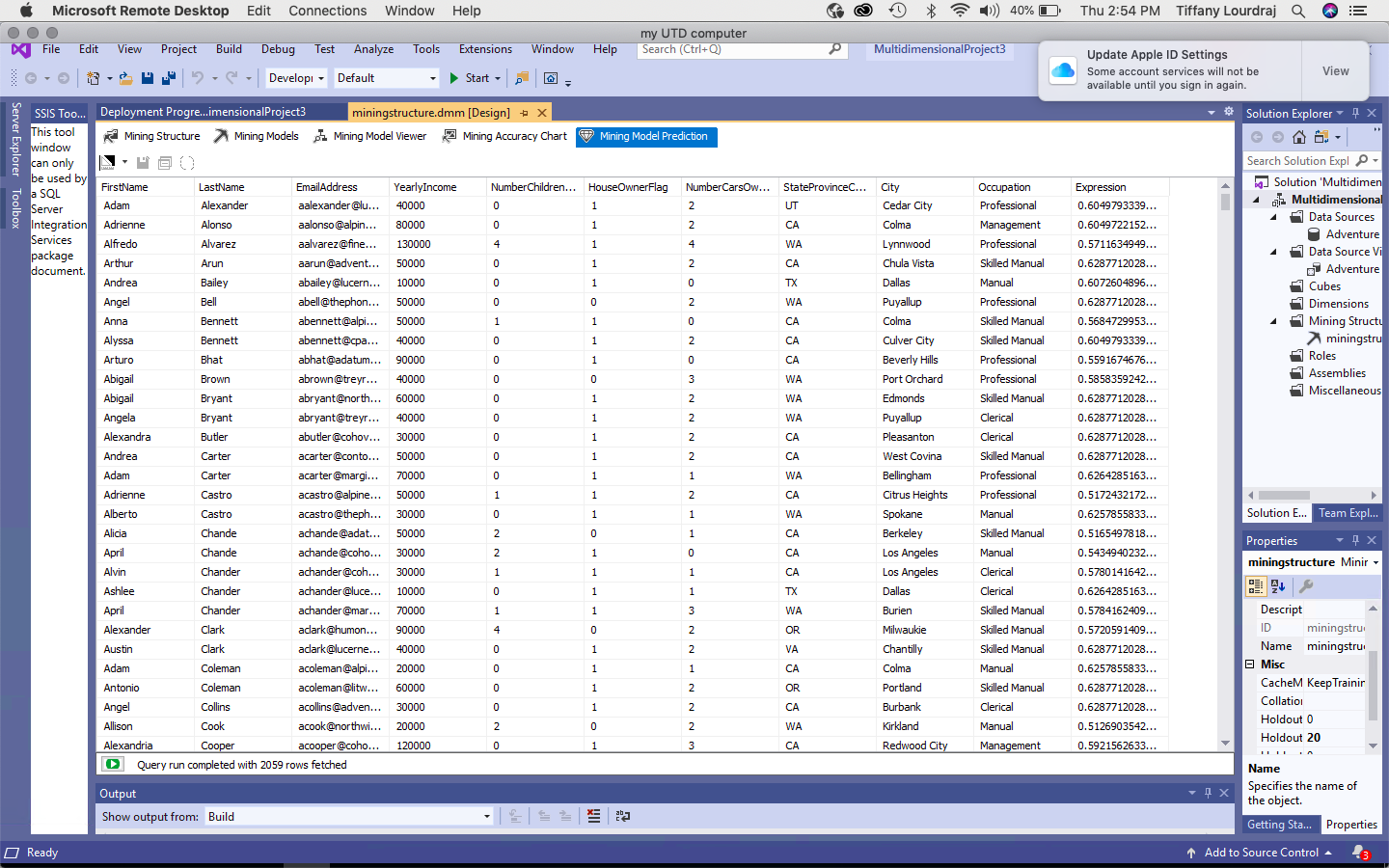
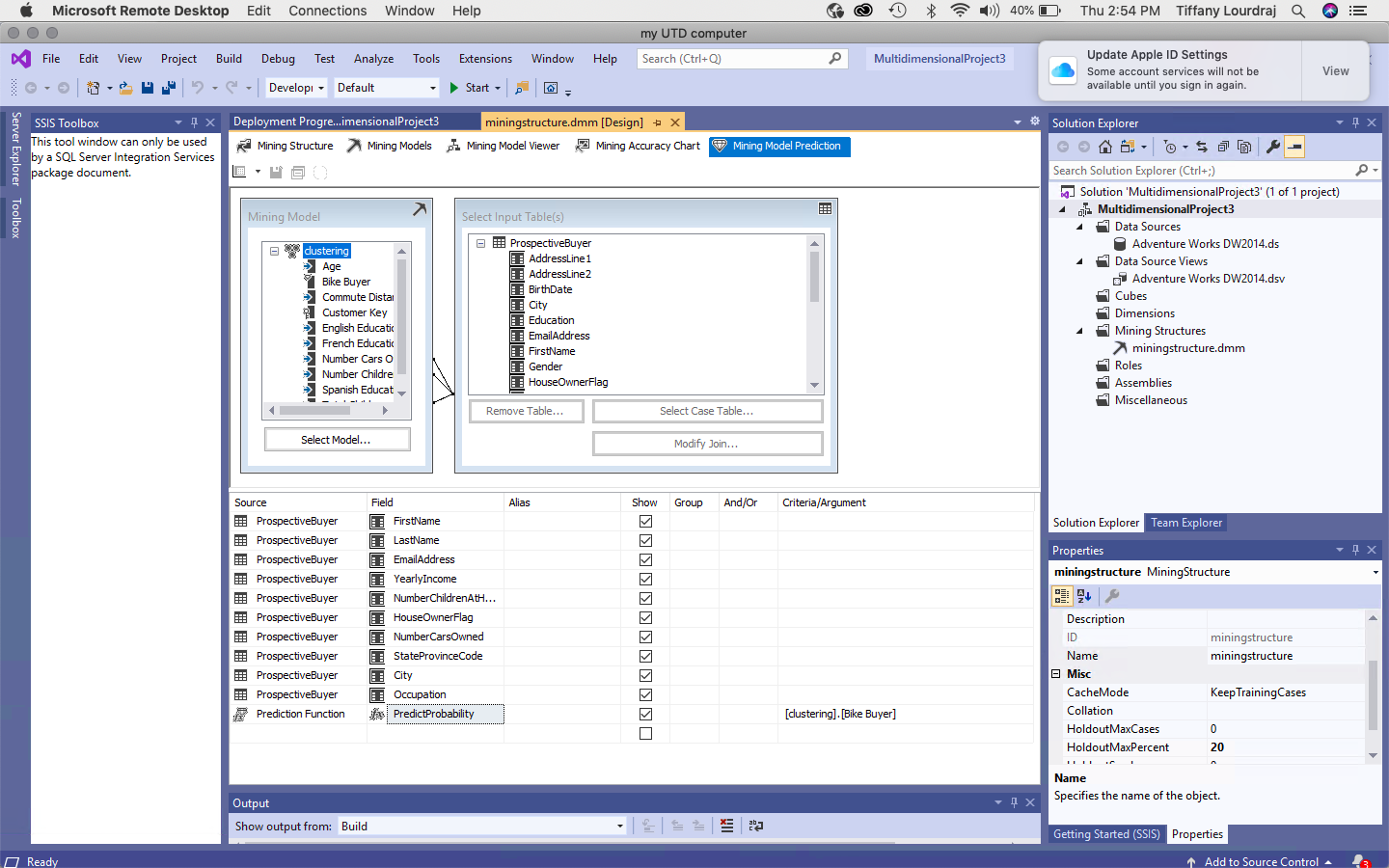
1. Compare Cluster 1 to Cluster 4. **Paste a screenshot of this comparison and describe your findings in at least 30 words – 20 points**

**I found that total children, number children at home, age and education all have a impact on the clusters. Total children have a correlation to bike buying because most families will resort to vans to travel together. In addition, the education affects cluster 4 and should be be referenced if wanted to find a correlation.**

****

1. Using Mining Model Prediction, find out the likelihood of someone buying a bike (Hint: Use the Predict Function to Predict Probability), using the Clustering Model and the ProspectiveBuyerTable. Include the following attributes
   1. Prospective Buyer Key
   2. First Name
   3. Last Name
   4. Email Address
   5. House Owner Flag
   6. StateProvinceCode
   7. City
   8. Occupation
   9. Yearly Income
   10. Number of Cars Owned
   11. Number of Children at Home

**Paste the screenshot of the results(whatever can be captured in one go) AND the actual query(not the screenshot of the query)– 20 points**



SELECT

t.[FirstName],

t.[LastName],

t.[EmailAddress],

t.[YearlyIncome],

t.[NumberChildrenAtHome],

t.[HouseOwnerFlag],

t.[NumberCarsOwned],

t.[StateProvinceCode],

t.[City],

t.[Occupation],

PredictProbability([clustering].[Bike Buyer])

From

[clustering]

PREDICTION JOIN

OPENQUERY([Adventure Works DW2014],

'SELECT

[FirstName],

[LastName],

[EmailAddress],

[YearlyIncome],

[NumberChildrenAtHome],

[HouseOwnerFlag],

[NumberCarsOwned],

[StateProvinceCode],

[City],

[Occupation],

[TotalChildren]

FROM

[dbo].[ProspectiveBuyer]

') AS t

ON

[clustering].[Total Children] = t.[TotalChildren] AND

[clustering].[Number Children At Home] = t.[NumberChildrenAtHome] AND

[clustering].[Number Cars Owned] = t.[NumberCarsOwned]