(Serial Number)

Project Log Sheet – Supervisory Session

Note on use of the project log sheet:

1. This log sheet is designed for all formal meetings, of which there must be at minimum SEVEN (7) during the course of the project (SEVEN mandatory supervisory sessions).
2. The student should prepare for the supervisory sessions by deciding which question(s) he or she needs to ask the supervisor and what progress has been made (if any) since the last session, and noting these in the relevant section of the form, effectively forming an agenda for the session.
3. A log sheet is to be brought by the STUDENT to each supervisory session.
4. The actions by the student (and, perhaps the supervisor), which should be carried out before the next session should be noted briefly in the relevant section of the form.
5. It is recommended that students bring along log sheets of previous meetings during each supervisory session.
6. The log sheet is NOT a deliverable for the project but it is an important record of a student's organization and learning experience. The students will be asked to hand in the log sheets as an appendix of the final report, with sheets dated and numbered consecutively. This is an important part of evidence on how you managed your project during the semester.

Student's Name: .....Linni Qin..... **Date:** …29/09/2017……..**Meeting No: 3**…

Project title: IT MASTERS (DS): Perform a Data Science Analysis of a Dataset/Task **UNIT:** ……IFN701……

☐ Journal entry logged into Blackboard (Optional)

Supervisor's Name: …… Guido Zuccon …… **Supervisor's Signature:** Guido Zuccon

Update on progress since last meeting, and challenges faced if any (noted by student before mandatory supervisory meeting):

1. **Transaction data and the property data have been merged**
2. **Missing data status have been identified**
3. **Correlation between log error and each property feature has been explored**
4. **Binary Prediction models were applied.**

Challenge:

1. **No correlation exists between the log error and the 57 property features**
2. **Binary prediction models do not fit Zillow's dataset**

Items for discussion (noted by student before mandatory supervisory meeting):

1. **Will the linear regression suit this project?**
2. **The way to identify the proper combination of training data and test data**
3. **The introduction about the cross-validation**
4. **Mean Squared Error is recommended to evaluate the prediction model**

Action List (to be attempted or completed by student by the next mandatory supervisory meeting):

1. **Use different groups of training data and test data to identify the best combination dataset for model training. For example, use the first three-week data to predict the last week value, use three-months data to predict the value for a future week.**
2. **Apply Mean Squared Error to evaluate the performance of the training model**
3. **Find other R techniques to identify the correlation between the log error and the variable by groups**