**Project diary** [Copy and paste the box for each week. No more than one A4-page per week.]

|  |  |
| --- | --- |
| **Week** | 1 (Week begin 30/03/2020) |
| **Communication** | * Fiona made original contact with the group (28/03/2020) * All but two in the group responded by 01/04/2020 * Amit suggested we set a time and date to discuss and divide the work and Fiona suggested we do this by the 01/04/2020. * Fergal and Fiona discussed their ideas and progress via Microsoft teams (04/04/2020). Vishal and Amit contacted the group later, with Vishal proposing work to be done in Tableau. |
| **Any analysis or other difficulties that arose and how they were overcome** | * Finding contact details of members of the group * Fergal had issues parsing unicode into R when loading as dplyr tibble, solution was to load into R as dataframe first and then parse. * Fiona had issues reading in the data due to euro symbols. Solution was to read in without header=TRUE, add new column headings to correspond with Fergal’s and removing the now defunct row 1 (the original headers). |
| **Progress made** | * Fergal set up two github repos and a project WhatsApp group. * Fergal pushed initial .csv data to repos, added basic parsing for Price column to convert to numeric and began investigating Choropleth Maps for representing data. * Fiona started looking into the individual towns in more detail. She decided on 16 towns across Ireland and extracted the data for these and made a preliminary start on visualisation of the data. * Vishal use tableau to make some initial visualization. |
| **Member contributions to progress** | * Choropleth Map code - Fergal * Town code – Fiona * Amit : In the tableau, created map for median price for each county and a graph for change in median price for the top 5 county over the years. |

|  |  |
| --- | --- |
| **Week** | 2 (Week begin 06/04/2020) |
| **Communication** | * Discussed with team members using Tableau as well for visualisation. * Meeting was held on the 11th April. Fiona suggested that we should all work off the one clean dataset, and offered to make a start on cleaning the data once this was agreed by the team members present. She also pointed out that the project description asks us to look at house price change over a range of Irish towns as opposed to counties. * A deadline of the 22nd May was set at the meeting to have a draft report complete. * A second meeting was held on 12th April. Ramesh suggested that he will schedule and organise the meetings. |
| **Any analysis or other difficulties that arose and how they were overcome** | * Rgeo package required rebooting rstudio in order to work. * Multi-unit complexes are recorded in quite a variety of ways. Fiona contacted the group to discuss how we were to deal with this and we decided to drop any properties over 10 million euro. Fergal came up with some code to split multi-property units into individual units – with an imputed mean house price. This was added to the code Fiona was working on to clean the dataset. * For the town dataset that Fiona was working on some of the house addresses are in Irish so fadas were throwing out error messages. Fiona used str\_replace\_all(hp$ address,"[^[:alnum:]]", " ") to get around this issue. |
| **Progress made** | * Fergal sourced .shp file to use for choropleth map, loaded .shp file into r and plotted basic map of Ireland (without analysis) * Fiona has looked at median house price change over time for 16 different towns around Ireland. * Fiona and Fergal worked on cleaning the dataset and uploaded it to github for the rest of the team to check on 12th April. * Vishal use tableau to visualize the average house price for different counties and use data to predict future price. * Vishal start working on average house price for different Irish towns. |
| **Member contributions to progress** | * Fiona: cleaned dataset with aid of Fergal. Produced code to extract data on 16 towns around Ireland to work with. * Vishal: Visualization on how average house price for different counties have been changed over time. * Amit: In the tableau, created map on the proportion of new property and property with full market price in different county. |
| **Week** | 3 (Week begin 13/04/2020) |
| **Communication** | * A large number of visualisations from members of the team were presented to the rest of the group via email. * Meeting held on 15th April and meeting minutes were sent by Ramesh. * We agreed that each member would have half a page to present their visualisations along with supporting test. The final half page would be used for our conclusions. |
| **Any analysis or other difficulties that arose and how they were overcome** | * Duplicate properties within the dataset. Some of these are due to the same property being sold multiple times throughout the timeframe, some are an artefact of splitting the multi-unit properties when cleaning the dataset and some are how they were entered onto the PRR to being with. For the last category, we have no way of knowing which are errors and which are multi-unit properties all sold under the one address. We therefore left them in. |
| **Progress made** | * All team members are firming up on their visualisations |
| **Member contributions to progress** | * John: heat-map displaying house price change for Dublin commuter belt * Fergal: Added county titles to choropleth map, updated maps to show median instead of mean and worked on overview section of report. * Ramesh: new house and second-hand house variation across Dublin and other towns * Fiona: visualisation on house price change over time for nine different regions in Ireland, based on four towns per region * Vishal: Visualization on average house price changing over time for different Irish towns and use data to predict house price for year 2020 and 2021. * Vishal compare the average house price and total sale in four major towns of Ireland and found that total sale is more in east of the country, possibly indicating the economic divide. * Amit: In the tableau, created a visualization to display the median price in top towns and its variation over the year, And a heat map to indicate change in median price in and around the region. |

|  |  |
| --- | --- |
| **Week** | 4 (Week begin 20/04/2020) |
| **Communication** | * Fergal communicated to the team that he had created a main .rmd file for the project, available on github * A meeting was held on 22nd April with full team attendance. We all presented our work and finalised which plots were to go into the report. * Fiona reminded the team that we need to work on the project diary too. * Another meeting was held on 25th April 2020. We discussed formatting and ways to tidy up the code (remove the cleansing data code and creation of towns dataset code from the main .rmd, justify text, use ‘property’ rather than ‘house’, refer to figures in text). Meeting minutes were sent by Ramesh for both meetings. |
| **Any analysis or other difficulties that arose and how they were overcome** | * Fiona had trouble wrapping text around her plot. Ramesh suggested using <div> </div> which worked perfectly. * Fiona moved the project diary from github to the project’s Microsoft Teams files folder to use as a shared document. This should be a much faster, more convenient way to make updates to the diary. |
| **Progress made** | * Fergal created a main .rmd file for the project and added his choropleth map plus text to it. * John added his code and heatmap to the main .rmd file. * Fiona added her code and regional towns plot to the main .rmd file. She also uploaded the town dataset and code, as well as the code for cleansing the data to github. * Vishal and Amit added a copy of their visualisations to the project’s Microsoft Teams group files folder. * Ramesh added his code and plot to the main .rmd file and made some formatting changes. |
| **Member contributions to progress** | * Fergal created the main .rmd file. * Fergal, John, Fiona and Ramesh all added their components to the main .rmd file. * All team members attended the meeting where we decided on which plots to present in the report. * Vishal and Amit added a copy of their visualisations to the project’s Microsoft Teams group files folder. |

|  |  |
| --- | --- |
| **Week** | 5 (Week begin 27/04/2020) |
| **Communication** | * Meeting on the 28th April 2020 to discuss formatting and final edits * Meeting on the 29th April 2020 to finalise the report. |
| **Any analysis or other difficulties that arose and how they were overcome** | * Fiona thought she had a bug in her code as it would run fine if in a separate file but not within the main .rmd file after an initial code run was made. Ramesh noticed that adding ‘plyr’ after ‘dplyr’ prevents some code acting as expected so he added code which detaches the ‘plyr’ and ‘dplyr’ before loading the libraries and also moved all libraries to the one place. This has resolved the ‘bug’ in Fiona’s code. * Fiona’s plot will not plot correctly if an older version of R is used. It will plot perfectly when code is run on the RStudio server. A recommendation to use RStudio server to run the code has been added to the readme file accompanying the project. |
| **Progress made** | * Fergal and Fiona uploaded their edits to the main .rmd file, and updated their supporting files too. * Fiona wrote the conclusions, added some methodology and did some minor editing to the report .rmd file. |
| **Member contributions to progress** | * All members of the team attended the meeting on 28th to go over final edits and formatting. * All members of the team will do a final check of the project folder before submission. |