



APPENDIX BB

Base Year and Projected Occupancy for WRMP19

ABSTRACT

Technical appendix for the approach to calculating the Unmeasured/Measured Split in household occupancy in the 2015/16 base year and the projection of occupancy out to 2044/45.

Contents

Overview of Approach	1
The customer base	1
Survey Response.....	2
Classification Model.....	3
Predicted Occupancies	3
Balancing Occupancy.....	7
Projected Occupancy.....	8

Overview of Approach

In March 2017 Portsmouth Water conducted an online survey of over 2,600 customers using a stratified sampling approach using known features of the customer base; demographic, meter status, flat status, and new/existing property status to create 128 customer types. Using the survey data a classification model was used to assign probabilities that a customer type belonged to an occupancy of 1 to 6. The implied probabilities are used to provide an estimation of the population for each of the 128 groups. The populations are aggregated up to simply two groups, Measured and Unmeasured. The raw result is an occupancy of 2.21 for a Measured property and 2.47 for an Unmeasured Property. There is a deficit of 22,455 persons (3%) compared to the company population for the 2015/16 base year which is provided by Experian requiring a rebalancing exercise. The rebalanced final occupancy is 2.27 for a Measured Property and 2.49 for an Unmeasured Property.

The customer base

A breakdown of the customer base according to the Portsmouth Water billing system as at March 2017 is presented below

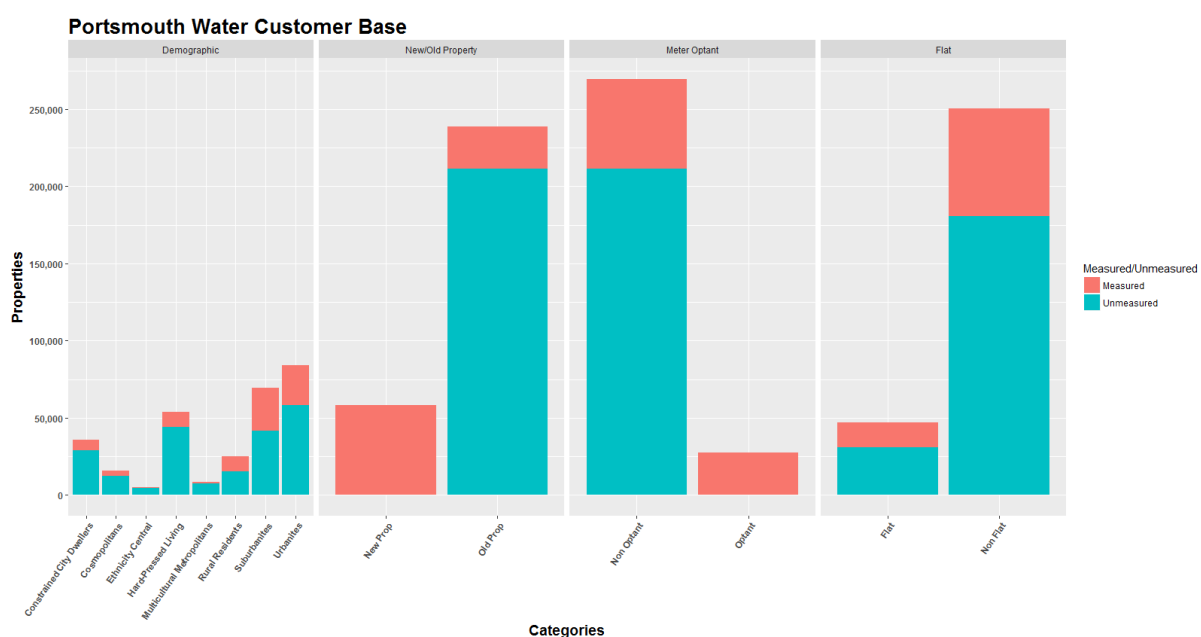


FIGURE 1: PORTSMOUTH WATER CUSTOMER BASE BREAKDOWN

Survey Response

In Spring 2017 Portsmouth Water commissioned a survey in order to obtain occupancy data receiving over 2,600 responses. After data cleansing the final data set contained 2,430 responses.

Figure 2 shows the breakdown of the 2,430 responses, the survey broadly reflects the customer base in

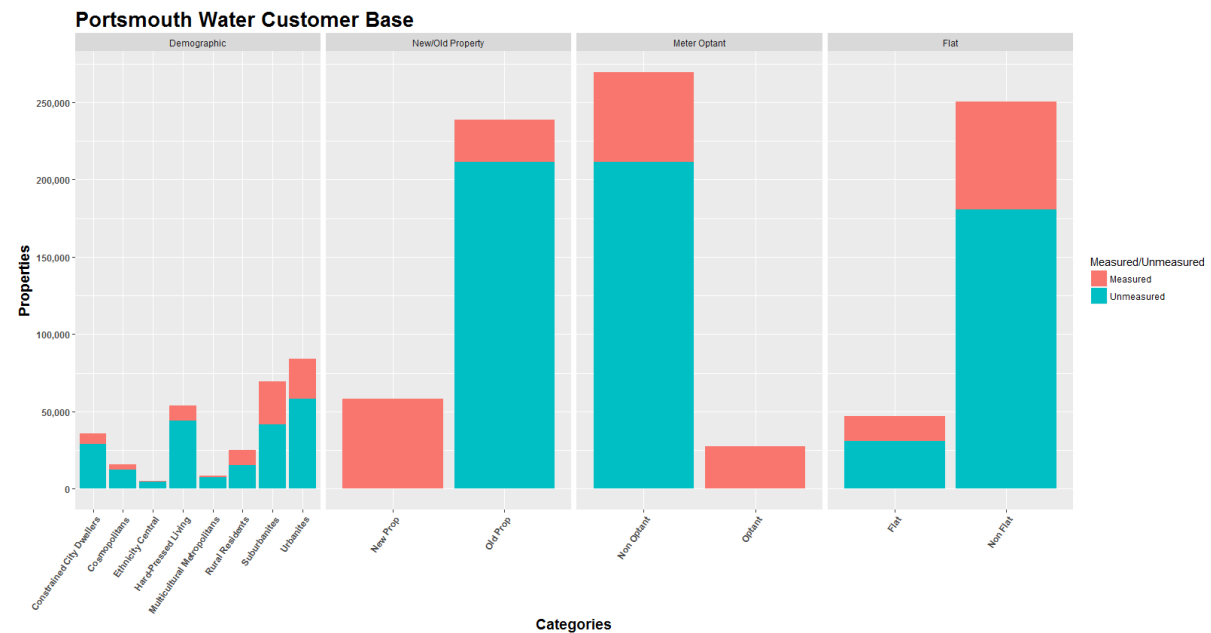


Figure 1

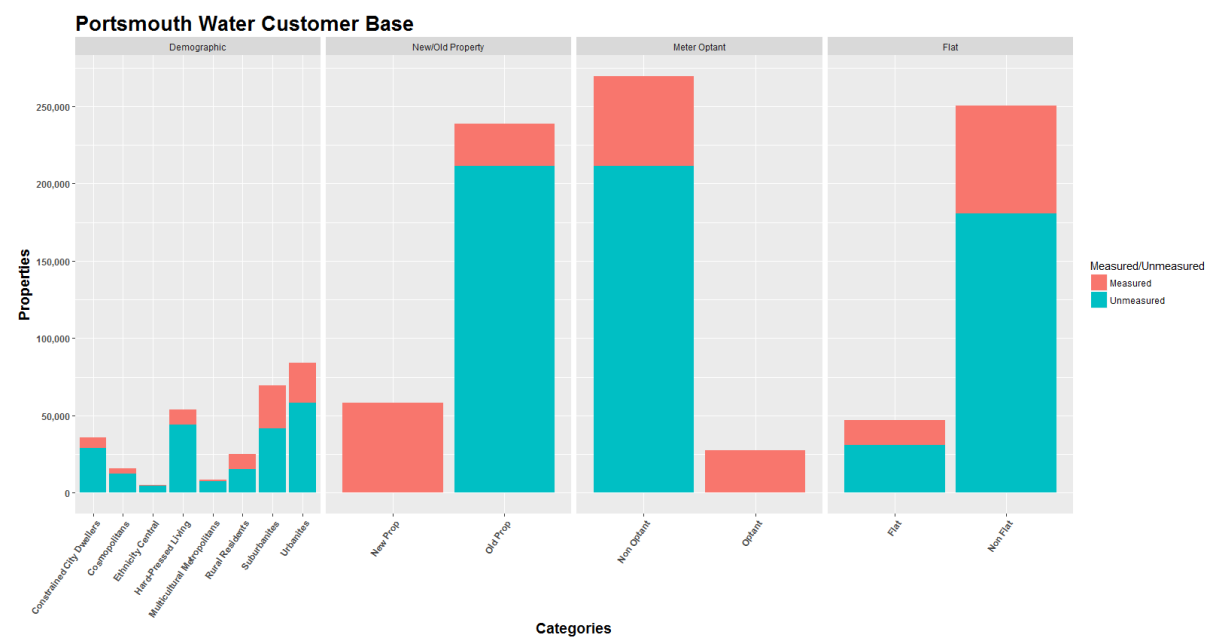


Figure 1

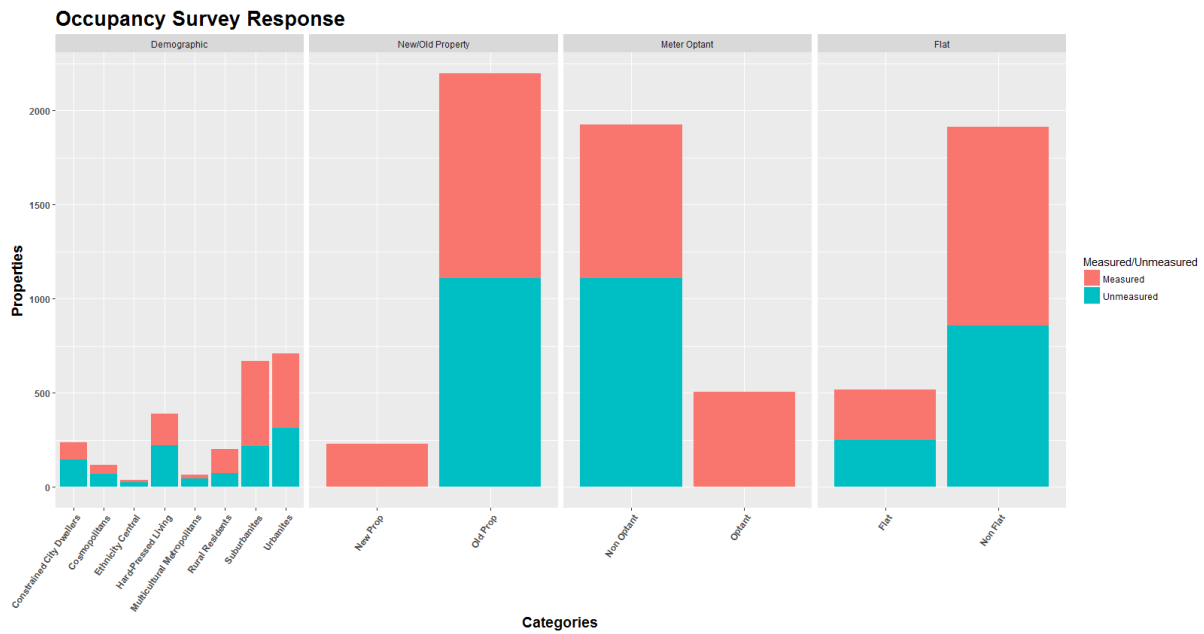


FIGURE 2: SURVEY RESPONSES BY CUSTOMER TYPE

In Figure 3 the percentage splits by customer type provide some insight into the difference in occupancy between the groups.

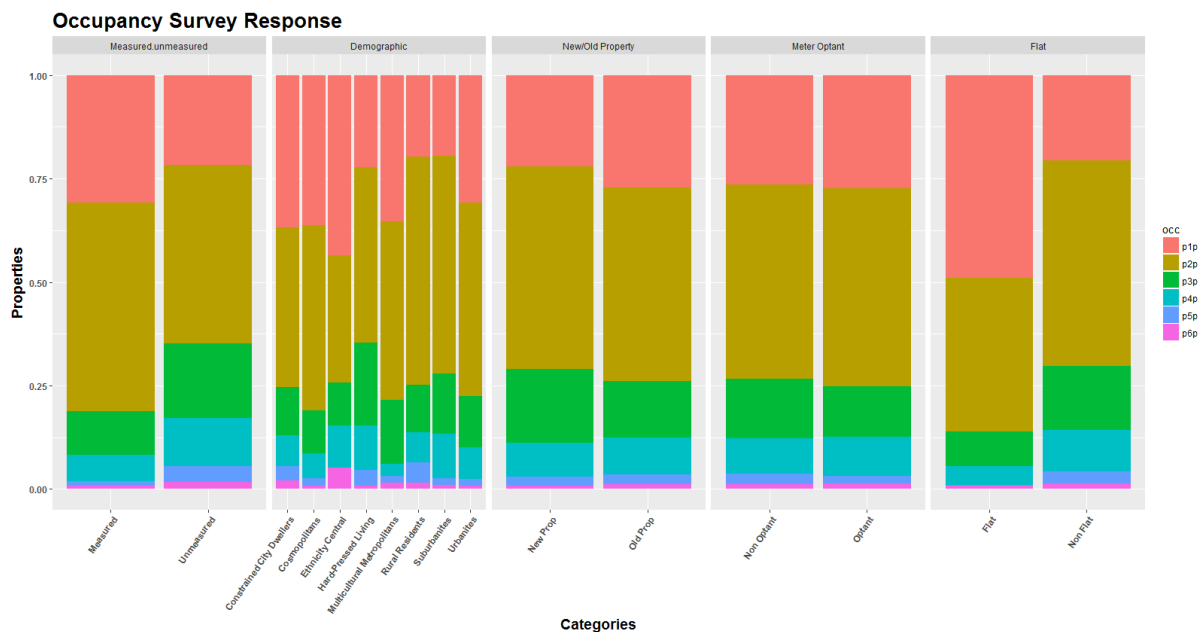


FIGURE 3 PERCENTAGE SPLITS OF SURVEY RESPONSE BY OCCUPANCY

The Measured/Unmeasured split of customers shows the Measured customers to have a significantly higher proportion of one and two occupancy properties compared to the Unmeasured group.

The split using ONS demographic types shows some differences between the groups. The 'Constrained City Dwellers', 'Cosmopolitans' and 'Ethnicity Central' groups tend to show a greater weighting to lower occupancies. The 'Hard-Pressed living', 'Rural Residents' and 'Suburbanites' tend to demonstrate a higher weighting to larger occupancies.

The splits by 'New/Old property' and 'Optant/Non-Optant' show similar percentage splits in occupancy.

The Flat/Non-Flat status shows unsurprisingly that the flats have a heavier weighting to single occupancy properties.

Classification Model

A classification model is used to provide predicted probabilities that each customer type will have an occupancy of 1 to 6. The purpose of the model is that it will allow for a generalised prediction to be made back against the known features of the customer base including combinations of factors that may have a low representation in survey sample.

A tree based Gradient Boosting approach is used to assign probabilities. Gradient Boosting is a machine learning technique that can be used for multi-classification problems such occupancy prediction. The approach uses an ensemble of decision trees to provide a likelihood of a customer type having an occupancy of 1 to 6.

Predicted Occupancies

The predicted probabilities ranked are presented in Table 1 .

By multiplying the occupancy by the weighted probability, a predicted occupancy is derived for each group, this figure is presented in the far right column of Table 1.

The predicted results are logical and also reflect the original survey data. Those groups with the lowest occupancy are flats and tend to be metered. Those with higher occupancies are unmeasured properties which are not flats and are more likely to be from demographic groups that are financially constrained or from demographic groups that live in more rural areas.

Supergroup Name	Measured Status	New Property	Flat	Meter Optant	No. Prop	1p	2p	3p	4p	5p	6p	Occ
Multicultural												
Metropolitans	TRUE	FALSE	TRUE	TRUE	85	0.64	0.27	0.05	0.02	0.00	0.01	1.51
Urbanites	TRUE	FALSE	TRUE	TRUE	1605	0.54	0.35	0.06	0.03	0.00	0.01	1.63
Constrained City Dwellers	TRUE	FALSE	TRUE	TRUE	550	0.53	0.36	0.05	0.04	0.01	0.01	1.68
Multicultural												
Metropolitans	TRUE	TRUE	TRUE	FALSE	438	0.52	0.32	0.13	0.02	0.00	0.01	1.69
Cosmopolitans	TRUE	FALSE	TRUE	TRUE	566	0.49	0.39	0.06	0.05	0.00	0.01	1.72
Urbanites	TRUE	TRUE	TRUE	FALSE	4675	0.44	0.42	0.10	0.03	0.00	0.00	1.76
Multicultural												
Metropolitans	FALSE	FALSE	TRUE	FALSE	1184	0.46	0.40	0.10	0.02	0.01	0.01	1.76
Rural Residents	TRUE	FALSE	TRUE	TRUE	150	0.49	0.38	0.06	0.05	0.01	0.02	1.76
Constrained City Dwellers	TRUE	TRUE	TRUE	FALSE	2179	0.43	0.40	0.14	0.03	0.00	0.01	1.80
Suburbanites	TRUE	TRUE	TRUE	FALSE	1646	0.38	0.47	0.11	0.04	0.00	0.00	1.81
Urbanites	FALSE	FALSE	TRUE	FALSE	8283	0.44	0.38	0.10	0.05	0.01	0.01	1.82
Constrained City Dwellers	FALSE	FALSE	TRUE	FALSE	8570	0.46	0.35	0.11	0.05	0.01	0.01	1.82
Suburbanites	TRUE	FALSE	TRUE	TRUE	447	0.41	0.46	0.05	0.06	0.00	0.01	1.83
Ethnicity Central	TRUE	FALSE	TRUE	TRUE	43	0.57	0.25	0.05	0.08	0.01	0.04	1.84
Rural Residents	TRUE	TRUE	TRUE	FALSE	574	0.43	0.36	0.17	0.03	0.01	0.01	1.84
Hard-Pressed Living	TRUE	TRUE	TRUE	FALSE	873	0.39	0.43	0.13	0.04	0.00	0.00	1.86
Suburbanites	FALSE	FALSE	TRUE	FALSE	1552	0.45	0.36	0.11	0.07	0.01	0.01	1.86
Rural Residents	FALSE	FALSE	TRUE	FALSE	702	0.42	0.41	0.10	0.04	0.02	0.01	1.87
Cosmopolitans	FALSE	FALSE	TRUE	FALSE	5783	0.42	0.38	0.12	0.06	0.01	0.01	1.89
Hard-Pressed Living	TRUE	FALSE	TRUE	TRUE	206	0.38	0.46	0.08	0.07	0.01	0.01	1.89
Cosmopolitans	TRUE	TRUE	TRUE	FALSE	1331	0.35	0.45	0.16	0.03	0.00	0.01	1.91
Multicultural												
Metropolitans	TRUE	FALSE	FALSE	TRUE	221	0.44	0.35	0.12	0.04	0.02	0.02	1.93
Hard-Pressed Living	FALSE	FALSE	TRUE	FALSE	2131	0.38	0.41	0.12	0.07	0.01	0.01	1.93
Ethnicity Central	TRUE	TRUE	TRUE	FALSE	380	0.39	0.34	0.18	0.06	0.00	0.02	2.02

Ethnicity Central	FALSE	FALSE	TRUE	FALSE	2767	0.49	0.25	0.11	0.11	0.01	0.04	2.03
Multicultural												
Metropolitans	FALSE	TRUE	TRUE	FALSE	1	0.30	0.43	0.22	0.03	0.01	0.02	2.08
Suburbanites	FALSE	TRUE	TRUE	FALSE	2	0.31	0.39	0.20	0.09	0.01	0.01	2.12
Urbanites	FALSE	TRUE	TRUE	FALSE	22	0.29	0.43	0.18	0.08	0.01	0.01	2.12
Urbanites	TRUE	FALSE	FALSE	TRUE	6472	0.27	0.48	0.14	0.08	0.02	0.02	2.13
Multicultural												
Metropolitans	TRUE	TRUE	FALSE	FALSE	445	0.27	0.44	0.22	0.04	0.02	0.01	2.14
Rural Residents	FALSE	TRUE	TRUE	FALSE	2	0.32	0.37	0.22	0.05	0.03	0.02	2.15
Constrained City Dwellers	FALSE	TRUE	TRUE	FALSE	3	0.30	0.37	0.25	0.06	0.01	0.02	2.15
Cosmopolitans	FALSE	TRUE	TRUE	FALSE	11	0.27	0.39	0.25	0.06	0.01	0.02	2.19
Hard-Pressed Living	FALSE	TRUE	TRUE	FALSE	1	0.25	0.42	0.23	0.08	0.01	0.01	2.20
Cosmopolitans	TRUE	FALSE	FALSE	TRUE	440	0.23	0.52	0.10	0.11	0.03	0.01	2.22
Urbanites	TRUE	TRUE	FALSE	FALSE	13216	0.16	0.57	0.16	0.09	0.02	0.01	2.25
Multicultural												
Metropolitans	FALSE	FALSE	FALSE	FALSE	6017	0.24	0.47	0.17	0.05	0.05	0.03	2.25
Constrained City Dwellers	TRUE	FALSE	FALSE	TRUE	1325	0.25	0.48	0.12	0.11	0.03	0.02	2.25
Suburbanites	TRUE	TRUE	FALSE	FALSE	16959	0.11	0.61	0.18	0.09	0.01	0.00	2.29
Rural Residents	TRUE	FALSE	FALSE	TRUE	3259	0.20	0.53	0.10	0.11	0.03	0.02	2.30
Constrained City Dwellers	TRUE	TRUE	FALSE	FALSE	2720	0.15	0.54	0.20	0.07	0.03	0.01	2.30
Cosmopolitans	TRUE	TRUE	FALSE	FALSE	868	0.12	0.59	0.18	0.07	0.02	0.01	2.31
Rural Residents	TRUE	TRUE	FALSE	FALSE	5796	0.14	0.55	0.20	0.06	0.04	0.01	2.34
Suburbanites	TRUE	FALSE	FALSE	TRUE	8887	0.16	0.55	0.13	0.13	0.01	0.02	2.35
Hard-Pressed Living	TRUE	TRUE	FALSE	FALSE	5889	0.14	0.48	0.25	0.09	0.03	0.00	2.39
Urbanites	FALSE	FALSE	FALSE	FALSE	49683	0.17	0.47	0.18	0.13	0.03	0.01	2.41
Ethnicity Central	TRUE	FALSE	FALSE	TRUE	60	0.32	0.34	0.08	0.16	0.03	0.07	2.44
Hard-Pressed Living	TRUE	FALSE	FALSE	TRUE	3116	0.17	0.45	0.21	0.13	0.03	0.01	2.44
Ethnicity Central	TRUE	TRUE	FALSE	FALSE	105	0.16	0.47	0.18	0.14	0.03	0.03	2.48
Constrained City Dwellers	FALSE	FALSE	FALSE	FALSE	20183	0.17	0.44	0.18	0.13	0.06	0.02	2.52
Cosmopolitans	FALSE	FALSE	FALSE	FALSE	6566	0.16	0.47	0.16	0.13	0.07	0.02	2.52
Rural Residents	FALSE	FALSE	FALSE	FALSE	14382	0.14	0.53	0.12	0.08	0.12	0.01	2.55

Suburbanites	FALSE	FALSE	FALSE	FALSE	40121	0.15	0.44	0.20	0.16	0.04	0.02	2.55
Hard-Pressed Living	FALSE	FALSE	FALSE	FALSE	41777	0.15	0.40	0.25	0.14	0.05	0.01	2.58
Multicultural												
Metropolitans	FALSE	TRUE	FALSE	FALSE	3	0.10	0.43	0.32	0.06	0.06	0.03	2.64
Urbanites	FALSE	TRUE	FALSE	FALSE	20	0.07	0.43	0.24	0.21	0.04	0.01	2.76
Ethnicity Central	FALSE	FALSE	FALSE	FALSE	1710	0.22	0.31	0.12	0.21	0.06	0.07	2.78
Suburbanites	FALSE	TRUE	FALSE	FALSE	14	0.06	0.37	0.30	0.20	0.04	0.01	2.83
Cosmopolitans	FALSE	TRUE	FALSE	FALSE	3	0.07	0.41	0.28	0.14	0.08	0.02	2.83
Constrained City Dwellers	FALSE	TRUE	FALSE	FALSE	3	0.07	0.38	0.31	0.15	0.07	0.02	2.84
Hard-Pressed Living	FALSE	TRUE	FALSE	FALSE	9	0.06	0.33	0.36	0.17	0.07	0.01	2.89
Rural Residents	FALSE	TRUE	FALSE	FALSE	10	0.07	0.41	0.23	0.11	0.17	0.02	2.95

TABLE 1: PREDICTED OCCUPANCY BY GROUP

Balancing Occupancy

The predicted occupancies are multiplied by the number of properties in each group to provide an estimation of populations in each group which can then be aggregated up to Measured/Unmeasured status.

	Properties	Estimated Population	Estimated Occupancy
Unmeasured	211,515	510,557	2.41
Measured	85,526	187,911	2.20

TABLE 2 MODELLED OCCUPANCY

The modelled occupancy from 2017/18 requires balancing with the base year 2015/16 outturn property figures from the billing system and the population figures provided by Experian for WRMP19/PR19. The modelled occupancy with 2015/16 figures leaves a residual population of 22,455, these are allocated proportionally to the Unmeasured and Measured customer according to the split of modelled population. The final rebalanced occupancies are 2.49 and 2.27 for Unmeasured and Measured properties respectively.

	2015/16 Properties	Modelled Occupancy	Modelled Population	Residual Population	Rebalanced Population	Rebalanced Occupancy
Unmeasured	210,156	2.41	506,476	16,745	523,221	2.49
Measured	78,509	2.20	172,720	5,710	178,430	2.27

TABLE 3 REBALANCED OCCUPANCY

Projected Occupancy

At a company level the projected occupancy is calculated using the Experian 'Trend Based' Population forecast with the Hybrid Forecast for New Properties in each year.

Measured and Unmeasured occupancy changes over time both as a result of changes in the company occupancy but also by the movement of unmeasured properties onto a meter. In order to reflect the movement of Unmeasured customers on to a meter, the company has a spreadsheet model which takes into account likelihood that lower occupancy properties are likely to switch first in addition to the introduction of new properties and populations.

To represent movement of customers from measured to measured status, the 62 customer types are aggregated into 4 cohorts based on the historic likelihood that the group will switch to a meter. This in turn dictates the change in occupancy over the 25 years for the Unmeasured and Measured groups, noting that the overall company occupancy is driven by the Experian population and property forecasts.

Cohort	Base Year Occupancy	No. Prop (2015/16)	Switch %
Most likely	1.92	40,380	39
Likely	2.40	52,537	30
Unlikely	2.61	96,935	28
Very Unlikely	3.27	20,304	03

TABLE 4 COHORTS OF UNMEASURED CUSTOMERS

The reduction in the unmeasured customer base over the 25 years is shown in Figure 4.

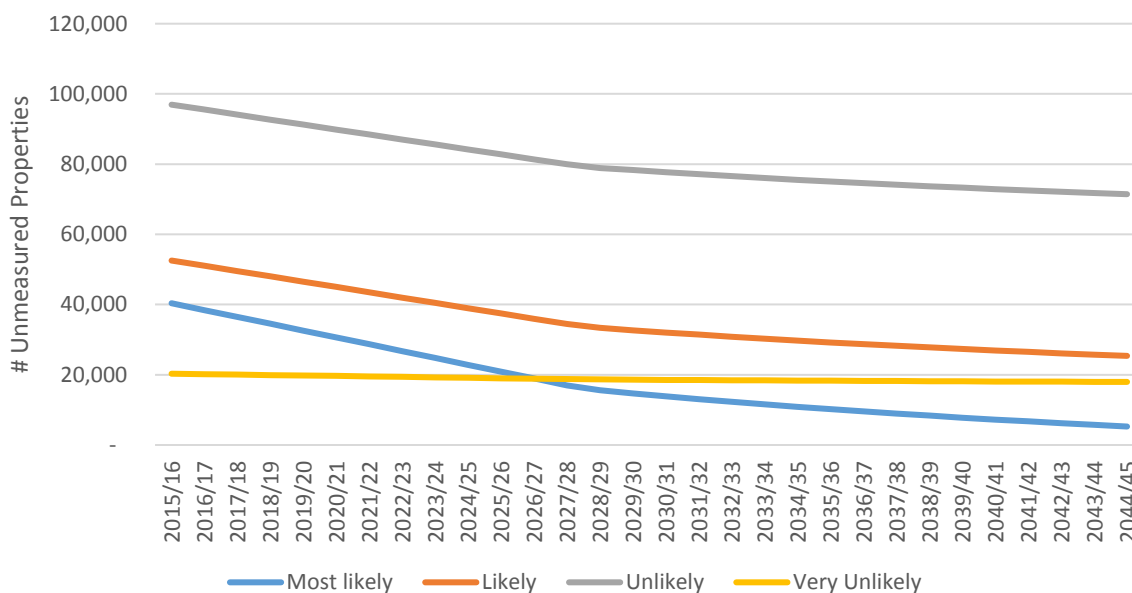


FIGURE 4 UNMEASURED PROPERTY NUMBERS OVER TIME

The resulting change in overall occupancy is presented in Figure 5.

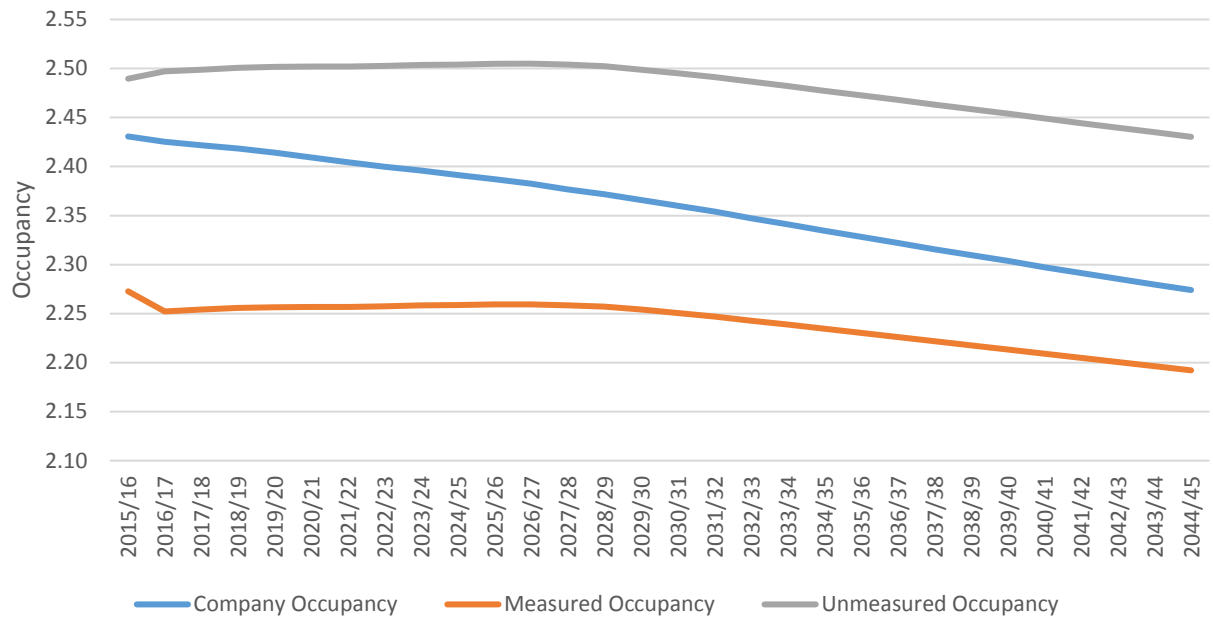


FIGURE 5 OCCUPANCY FORECAST

The unmeasured and measured occupancy remains flat up to 2029-2030. This is a result of 5,000 unmeasured customers a year moving on to a meter. The movement of lower occupancy properties from the unmeasured customer base to the measured customer base results in the unmeasured occupancy being propped up as higher occupancy properties are left in the customer base despite overall falling company occupancy. The measured occupancy remains level as a result of the introduction of meter optants with occupancies slightly higher than the existing customer base. From 2029-2030 the rate of movement from the unmeasured to the measured base has slowed resulting in both unmeasured and measured occupancy falling almost proportionally with the company level occupancy.