Millennial Patterns of Travel and Location Choice: The Same Old Hype or a True Change?

Jason Hawkins
PhD Candidate
Department of Civil Engineering
University of Toronto

Introduction

Popular culture is full of descriptions of the revolutionary characteristics of Millennials. Some of what is said about this generation is likely true, such as an increasing trend among men of staying home with the children (Livingston 2018). However, this is likely part of the longer-run pattern of increasing equality in the allocation of household roles, rather than a unique characteristic of this generation. One 2013 Time cover story describes the generation as "lazy, entitled narcissists who still live with their parents" (Stein 2013). Another article even outlines how Millennials are killing the mayonnaise industry (Hingston 2018). These sorts of characterizations are nothing new. One generation inevitably argues that the next generation is lazy and has it much better (although, is this a bad pattern?). It is not improbable that the ancient Greek philosopher Socrates complained about the slothful ways of his student Plato.

An issue that has been poorly defined, even in the academic literature, is that of the travel and location choice preferences of Millennials. This was a question that arose during workshop discussions at the most recent IATBR conference in July, 2018 and is the impetuous for the present work. Like all generations, Millennials are a heterogeneous group, and being the youngest adult generation research has matured in lock-step with the maturation of its members. As such, it is not surprising that choice patterns observed among Millennials in a 2012 study should disagree with those of a 2018 study. I hypothesize that some part of the research finding lower interest in car and home ownership (Klein and Smart 2017; Welke and McLaughlin 2015; Zhong and Lee 2017) is the result of this generation delaying life transitions relative to previous generations. For example, rather than marrying in their early 20s, many Millennials are choosing to delay this transition into their 30s. It is only with the most recent iteration of the national household travel survey (NHTS) that Millennials have aged sufficiently to manifest these life transitions.

I examine the hypothesis of life stage transition playing a stronger role than inherent generational differences through a series of ANOVA analyses. Drawing out a series of representative variables from the NHTS, I perform 2-way ANOVA tests using life stage (as defined in the NHTS) and generation (as defined by age). Separate tests are performed for each of full-time workers, part-time workers, and non-workers. I argue that travel and location choices are dominated by life stage rather than implicit characteristics of the generation. For example, a couple in their 20s with two children will be similar to a couple in their 40s with two children in travel and location choices, rather than a single person of a similar age.

Methodology

Analysis is based on the *perpub* table of the NHTS, which contains a wide range of statistics on the sociodemographic, transportation, and land use characteristics of 264,234 persons. I begin by cleaning the dataset to provide consistent estimation of regression models and ANOVA statistics (outlined in accompanying SQL file). The variable *R*_AGE is translated into generational bins according to definitions provided by the Pew Research Center (2018). Several categorical variables are assumed continuous to allow for continuous regression models to be estimated against them. This is a generally reasonable assumption. The variable *EDUC*, which gives a measure of the highest level of educational attainment by each respondent, is monotonically increasing in its numeric values for increasing levels of education. The variable *HOMEOWN* is of interest in this study because it helps to answer questions about the home ownership of Millennials. It is transformed into a binary (0/1) variable representing whether the respondent owns their dwelling. A continuous measure can be estimated for this variable by binomial logistic regression. However, ANOVA analysis cannot be performed on the resulting probability measure because it is purely a function of the factors defined in the ANOVA.

The main objective of this study is to quantify the variation in preferences for transportation and land use decisions by two demographic categorizations: generation and life stage. The principle method of analysis is ANOVA because it provides a straight-forward means of comparing differences between groups. Each of the variables of interest represents a *factor* with discrete categories within each *factor* defined as *levels.* Initial analysis used a 3-factor design, which included the additional variable of *WKFTPT* – whether the respondent works full-time, part-time or not at all. However, this introduced unnecessary complexity in model estimation and interpretation and the final model includes only the two *factors* of generation and life stage. To include differences by work status, separate analyses were performed for each of the *WKFTPT levels.* Life stage is defined according to the following levels:

1. One adult, no children
2. 2+ adults, no children
3. One adult, youngest child 0-5
4. 2+ adults, youngest child 0-5
5. One adult, youngest child 6-15
6. 2+ adults, youngest child 6-15
7. One adult, youngest child 16-21
8. 2+ adults, youngest child 16-21
9. One adult, retired, no children
10. 2+ adults, retired, no children

The ANOVA analysis is complicated by the presence of unequal *levels* between *factors* and unequal samples for each cell – defined by the number of respondents classified into each combination of generation and life stage pairing. The unbalanced nature of the ANOVA table means that it is non-orthogonal (i.e. $SSR \neq SS_A + SS_B + SS_{AB}$) and less efficient (i.e. larger parameter variances). The General Linear Model (GLM) F test statistic must be used, defined by (Helwig 2017):

$$\mathrm{F} = \frac{\mathrm{SSE_R} - \mathrm{SSE_F}}{\mathrm{df_R} - \mathrm{df_F}} \frac{\mathrm{df_F}}{\mathrm{SSE_F}}$$

where df_R anddf_F are the degrees of freedom for the reduced and full regression, respectively. Similarly, SSE_R and SSE_F are the sum of squared errors for the reduced and full regression residuals, respectively. There are several ways to define the reduced and full models and I employed 3 in the present analysis as follows (Helwig 2017):

1. Type I: Sequential addition of variables. Added explained variation by the inclusion of the *factor* variable of interest from the model design (e.g. Adding *factor* A)
2. Type II: Exclusion of variables. Reduced explained variation by the exclusion of all variables containing the *factor* of interest from the model design (e.g. A + B + AB vs. B only)
3. Type III: Partial sum of squares. Reduced explained variation by the exclusion of the *factor* variable of interest from the model design (e.g. A + B + AB vs. B + AB)

Each method gives a slightly different result, but all are founded on standard methods of comparing F-statistics against a test value (i.e. $\alpha = 0.05$).

As a means of adjusting for in-sample heteroskedasticity and low parameter efficiency, I test several forms of heteroskedasticity-consistent covariance estimators proposed by MacKinnon and White (1985). These methods amount to estimating some form of heteroskedastic random error term, which typically follows a normal distribution. MacKinnon and White propose 3 such adjustment terms, which they denoted by HC_1, HC_2, and HC_3, respectively. The robustness of the adjustment term increases from HC_1 to HC_3.

In additional to determining whether a *factor* is significant in its explanation of variation, I also calculate its size. Effect size estimation is more common in psychology and controlled experiments but proved useful in this context. I use the η^2 statistic developed by Cohen and Friedman (1973), defined by:

$$\eta^2 = \frac{df_A F_A}{df_A F_A + df_E}$$

where df_A is the degrees of freedom for the *factor* of interest, F_A is its associated F-statistic, and df_E is the degrees of freedom for the error term (total observations less the degrees of freedom for all *factors* and *interaction* terms).

Results

I consider the effect of life stage and generation on a set of 11 variables: home ownership, bikeshare use, rideshare use, carshare use, highest level of education attainment, vehicles per licenced household driver, public transit use in the past month, population of urban area of home location, frequency of working from home in the past month, thousands of miles driven in the past year, commuting time in minutes, and number of activity trips made in the past week. ANOVA tests were performed for each variable, except home ownership for which a logistic regression was completed. A summary of results using type I ANOVA is presented for non-worker respondent in Table A1. Similar tables were completed for each of part-time and full-time worker respondents and type II and type III ANOVA (provided in accompanying summary outputs). In nearly all instances, the two *factor* variables provide statistically significant descriptions of dependent variable variation. According to Table A1, life stage does not provide a significant explanation for variation in the frequency of working at home and neither *factor* variable provides a significant explanation for commuting time. However, this is to be expected given that the analysis is for non-workers and each is significant for part-time and full-time workers.

Based only on F-statistics, the ANOVA tests suggest that both life stage and generation provide significant explanations for each of the examined dependent variables. This does not give a conclusive result as to the absolute and relative effects of each *factor* variable. Cohen (1973) recommends a threshold of 0.02 for the η^2 statistic indicating a small, but significant, effect size. Calculating the η^2 statistic for each indicates that the effect, in relation to the remaining unexplained variation, is insignificant in almost all cases ($\eta^2 < 0.02$). Results suggest that life stage has a highly significant effect ($\eta^2 > 0.26$) on the use of car sharing programs among both part-time and full-time workers. Similarly, bike share use is moderately affected by life cycle stage ($0.13 < \eta^2 < 0.26$).

We know that the explained variations are significant, according to the F-statistics, but that the effect sizes are small. Given these inconclusive results, I suggest a comparison of effect sizes between the two *factor* variables. A summary of this analysis is presented in Table 1 below. I take the ratio of effect sizes between life stage and generation for each of the dependent variables, which gives some indication of their relative effect sizes. This suggests that the effect size is larger for life stage than generation for most of the dependent variables (for all ANOVA test types). Many of these ratios are much larger than 1. For example, for full-time workers, the effect size for life stage on frequency of working at home is 4.7× the magnitude of generational effect size.

Table 1 Summary of Relative Effects for Each Demographic Segment

Demographic Segment	**ANOVA Type**	**Relative Effect Size** $\left(\# \frac{\eta 2 \text{ LIF\_CYC}}{\eta 2 \text{ R\_AGE\_CAT}} > 1\right)$	**Relative Effect Size (% of Total)**
Non-worker	I	6	54.6
	II	6	54.6
	III	10	90.9
Part-time worker	I	7	63.6
	II	8	72.7
	III	10	90.9
Full-time worker	I	7	63.6
	II	8	72.3
	II	10	90.1

Looking at home ownership, Figures A1 to A3 present plots of parameter estimates for each demographic segment. The results of the logistic regression indicate that home ownership tends to increase with age, but no conclusion can be drawn as to if this is a generational effect or simply a function of life stage and income. The parameters associated with life stage exhibit greater variation and this *factor* gives a stronger indication of market segmentation in the residential real estate market.

Conclusions

The NHTS allows the analyst to consider a wide variety of questions and provide a rich set of explanatory variables. In this study, I considered the question of transportation and land use choice as segmented by life stage and generation. A series of ANOVA analyses were conducted, and effect sizes compared between the two *factor* variables. The question of generational and life stage effects is a difficult one. It is challenging to disentangle their independent effects on each decision. What is clear is that we must maintain caution in ascribing changing patterns of travel and location choice to purely generational differences. The overall preferences of the population will change with time as a function of evolving gender roles and cultural norms, new technologies, and the effects of climate change and other systemic changes on patterns of travel and urban development. I find that the role of life stage on choice is similar, if not greater, than that of generation. This supports the development of activity-based travel demand models that consider the role of household composition and life stage on patterns of travel and location choice (Beige and Axhausen 2012; Miller et al. 2015).

The analysis presented in this report should be viewed as a first approximation to the question. There are several ways in which it could be improved upon to better draw out the relative influence of generation and life stage. A major shortcoming in the current framing of the problem is that I restrict analysis to the 2017 NHTS. This means that generational effects are confounded with the effects of age, which will also affect preferences. A similar set of ANOVA analyses could be performed for past surveys, with comparisons made between similarly aged survey respondents in each NHTS. We would need to ensure that a similar set of variables is available in each of the past surveys, which is unlikely to be the case in the earliest surveys from 1983 and 1990.

References

Beige, Sigrun, and Kay W. Axhausen. 2012. "Interdependencies between Turning Points in Life and Long-Term Mobility Decisions." *Transportation* 39(4): 857–72. http://link.springer.com/10.1007/s11116-012-9404-y (September 26, 2018).

Cohen, Jacob. 1973. "Eta Squared and Partial Eta Squared in Fixed ANOVA Designs." *Educational and Psychological Measurement* 33: 107–12. https://journals-scholarsportal-info.myaccess.library.utoronto.ca/pdf/00131644/v33i0001/107_eapeiffad.xml (September 25, 2018).

Helwig, Nathaniel E. 2017. *Factorial and Unbalanced Analysis of Variance*. http://users.stat.umn.edu/~helwig/notes/aov2-Notes.pdf (September 25, 2018).

Hingston, Sandy. 2018. "How Millennials Killed the Mayonnaise Industry." *Philadelphia Magazine*. https://www.phillymag.com/articles/2018/08/11/mayonnaise-industry-millennials/ (September 24, 2018).

Klein, Nicholas J, and Michael J Smart. 2017. "Millennials and Car Ownership: Less Money, Fewer Cars." *Transport Policy* 53: 20–29. http://dx.doi.org/10.1016/j.tranpol.2016.08.010 (September 26, 2018).

Livingston, Gretchen. 2018. *Stay-at-Home Moms and Dads Account for about One-in-Five U.S. Parents*. http://www.pewresearch.org/fact-tank/2018/09/24/stay-at-home-moms-and-dads-account-for-about-one-in-five-u-s-parents/ (September 26, 2018).

MacKinnon, Jg, and Halbert White. 1985. "Some Heteroskedasticity Consistent Covariance Matrix Estimators with Improved Finite Sample Properties." *Journal of Econometrics* (537). https://pdfs.semanticscholar.org/0dec/28fb8e3e2ea6fa68b26e1c920ca9673fa2a9.pdf (September 25, 2018).

Miller, Eric J, James Vaughan, David King, and Matthew Austin MASc Candidate. 2015. "Implementation of A 'next Generation' activity-Based Travel Demand Model: The Toronto Case." In *Conference of the Transportation Association of Canada*, Charlottetown, PEI. http://conf.tac-atc.ca/english/annualconference/tac2015/s9/miller.pdf (September 26, 2018).

Pew Research Center. 2018. "Millennials." http://www.pewresearch.org/topics/millennials/ (September 25, 2018).

Stein, Joel. 2013. "Millennials: The Me Me Me Generation." In *Time*, , 1–9. http://time.com/247/millennials-the-me-me-me-generation/ (September 24, 2018).

Welke, Gerd M, and Eric J McLaughlin. 2015. "Perceptions of Home Acquisition Priorities by Millennials: The 'entitled Generation.'" *International Research Journal of Applied Finance* 6(8): 505–14.

Zhong, Laitian, and Bumsoo Lee. 2017. "Carless or Car Later?" *Transportation Research Record: Journal of the Transportation Research Board* 2664: 69–78. http://trrjournalonline.trb.org/doi/10.3141/2664-08 (September 26, 2018).

Data Appendix

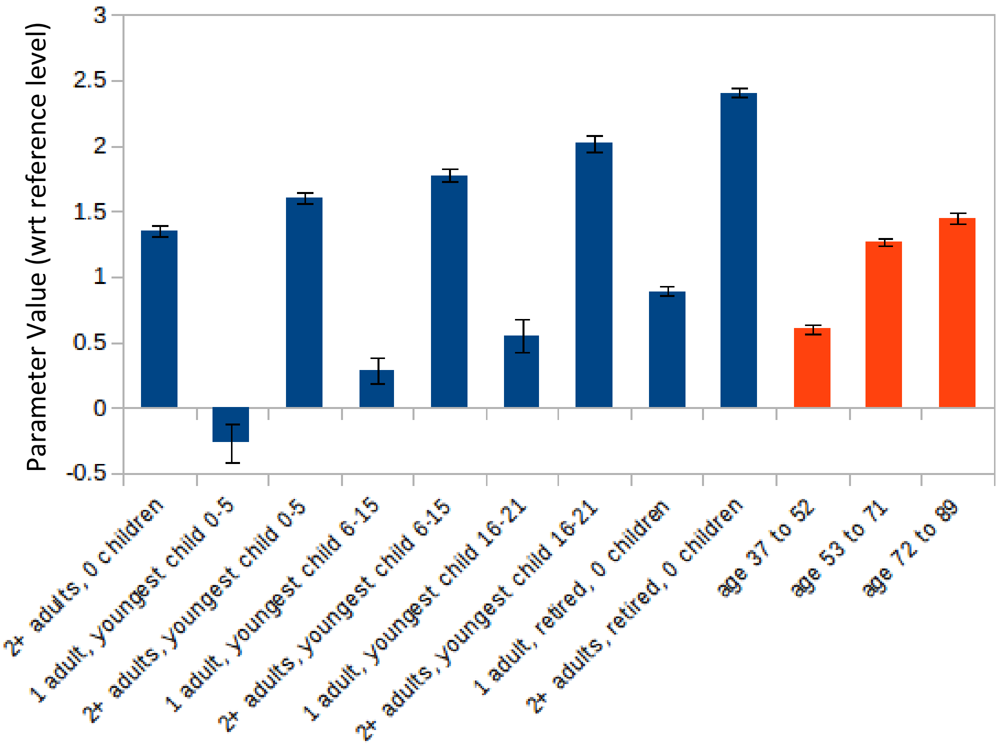

Figure A1 *Home Ownership Logistic Regression Results for Non-Workers*

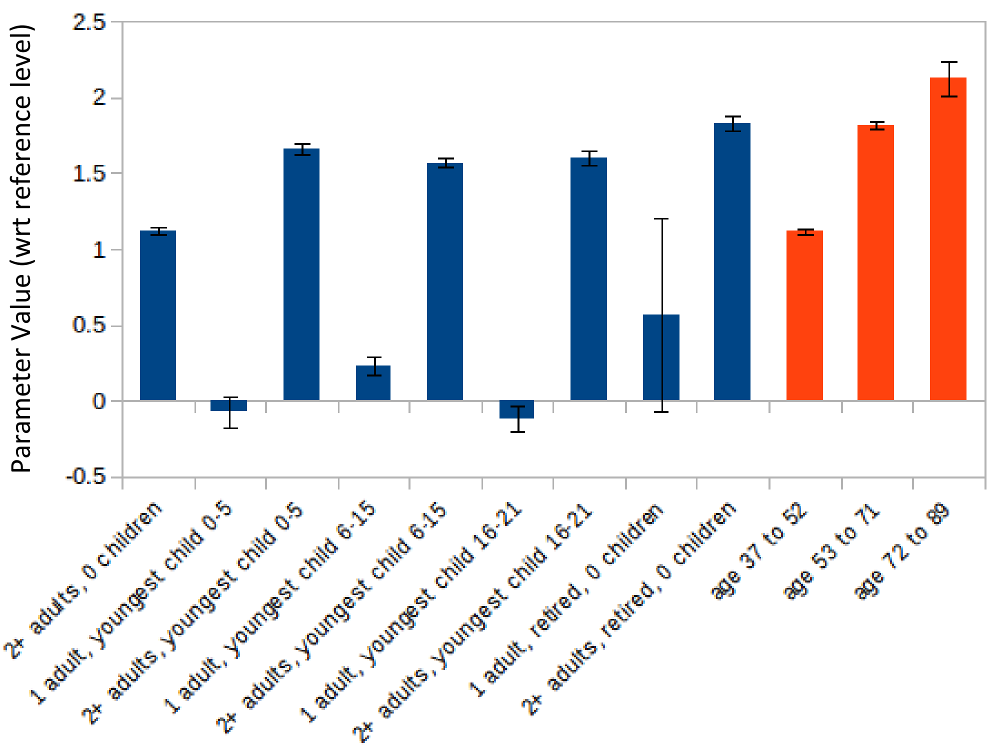

Figure A2 Home Ownership Logistic Regression Results for Part-Time Workers

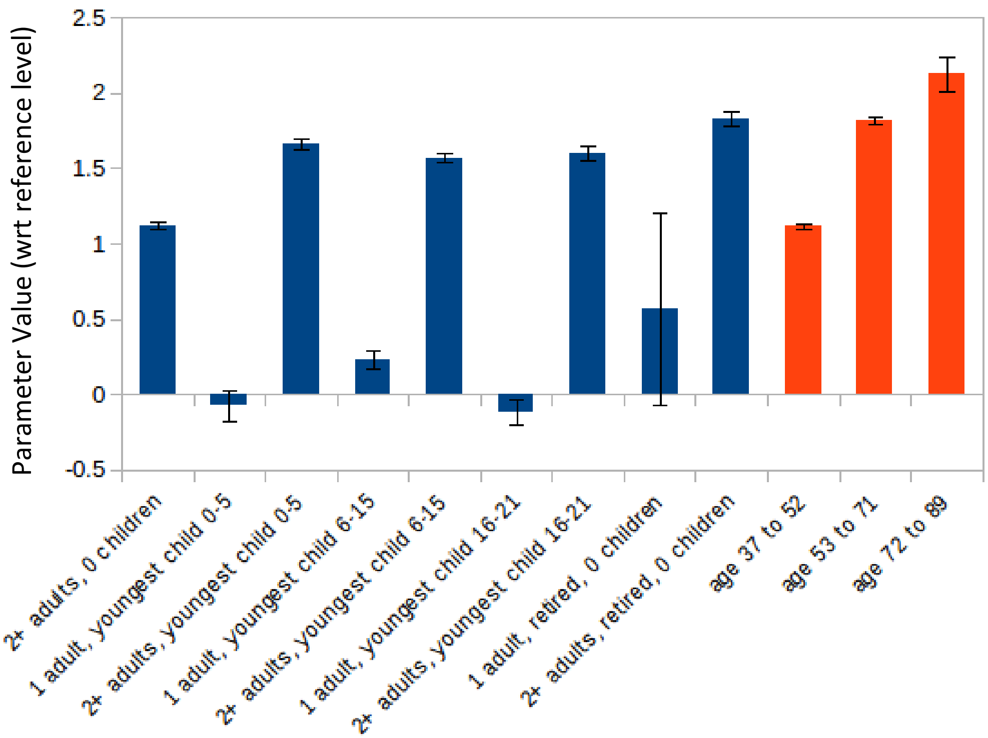

Figure A3 *Home Ownership Logistic Regression Results for Full-Time Workers*

Table A2 Summary of ANOVA Results for Non-Workers (Type I ANOVA)

	df	SS	F	PR(>F)	η^2	Ratio $\eta^2 \left(\frac{\text{LIF\_CYC}}{\text{R\_AGE\_CAT}}\right)$
BIKESHARE						
LIF_CYC	9	10.59	2.95	0.00	0.00	1.46
R_AGE_CAT	3	7.24	6.05	0.00	0.00	
LIF_CYC:R_AGE_CAT	27	13.39	1.24	0.18	0.00	
Residual	97700	38969.48				
RIDESHARE						
LIF_CYC	9	1094.23	95.00	0.00	0.01	1.03
R_AGE_CAT	3	1059.88	276.06	0.00	0.01	
LIF_CYC:R_AGE_CAT	27	742.36	21.48	0.00	0.01	
Residual	97700	125034.39				
CARSHARE						
LIF_CYC	9	12.34	5.22	0.00	0.00	0.90
R_AGE_CAT	3	13.72	17.42	0.00	0.00	
LIF_CYC:R_AGE_CAT	27	6.30	0.89	0.63	0.00	
Residual	97700	25652.01				
EDUC						
LIF_CYC	9	677.95	55.72	0.00	0.01	0.75
R_AGE_CAT	3	909.19	224.19	0.00	0.01	
LIF_CYC:R_AGE_CAT	27	1844.62	50.54	0.00	0.01	
Residual	97700	132073.05				
VEH_PER_DRIVE						
LIF_CYC	9	344.59	118.27	0.00	0.01	1.92
R_AGE_CAT	3	178.81	184.12	0.00	0.01	
LIF_CYC:R_AGE_CAT	27	49.17	5.63	0.00	0.00	
Residual	93871	30387.58				
PTUSED						
LIF_CYC	9	29346.53	282.06	0.00	0.03	5.39
R_AGE_CAT	3	5325.45	153.55	0.00	0.01	
LIF_CYC:R_AGE_CAT	27	3796.00	12.16	0.00	0.00	
Residual	97700	1129470.00				
URBANSIZE						
LIF_CYC	9	3242.43	107.88	0.00	0.01	6.86
R_AGE_CAT	3	468.39	46.75	0.00	0.00	
LIF_CYC:R_AGE_CAT	27	791.29	8.78	0.00	0.00	
Residual	97700	326282.98				
WKFMHMXX						
LIF_CYC	9	0.01	0.90	0.52	0.00	0.49
R_AGE_CAT	3	0.02	5.56	0.00	0.00	
LIF_CYC:R_AGE_CAT	27	0.04	1.39	0.09	0.00	
Residual	97700	93.94				
YEARMILE (x1000)						
LIF_CYC	9	10059.20	17.59	0.00	0.00	0.22

R_AGE_CAT	3	45245.36	237.33	0.00	0.01	
LIF_CYC:R_AGE_CAT	27	23183.34	13.51	0.00	0.00	
Residual	97700	6208486.00				
TIMETOWK						
LIF_CYC	9	1.17	1.54	0.13	0.00	3.09
R_AGE_CAT	3	0.38	1.50	0.21	0.00	
LIF_CYC:R_AGE_CAT	27	3.47	1.53	0.04	0.00	
Residual	97700	8224.07				
NACTTRP						
LIF_CYC	9	16033.74	24.47	0.00	0.00	0.69
R_AGE_CAT	3	23301.62	106.69	0.00	0.00	
LIF_CYC:R_AGE_CAT	27	7764.85	3.95	0.00	0.00	
Residual	97700	7112588.00				