

TheAnalyticsTeam

Sprocket Central Pty Ltd

Data analytics approach

[Peiqing Lian]

Agenda

1. Introduction
2. Data Exploration
3. Model Development
4. Interpretation

Customers Exploratory Analysis by Segmentation

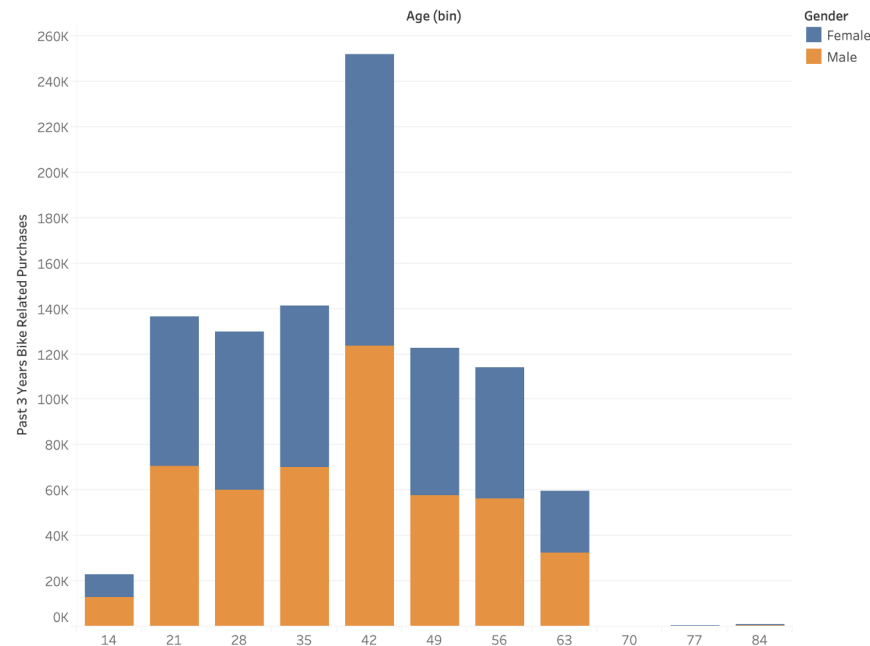
- Purchase behavior by Age & Gender
- Top Selling Products (Class & Line)
- Purchase behavior in Job Industry
- Purchase behavior by Brand
- Seasonality

Data Exploration

Purchase Behavior by Age & Gender

- **No significant distinction** among gender in terms of purchase
- Fewer purchase happens below age **20** and above **63**

Age&Gender Distribution by Purchase



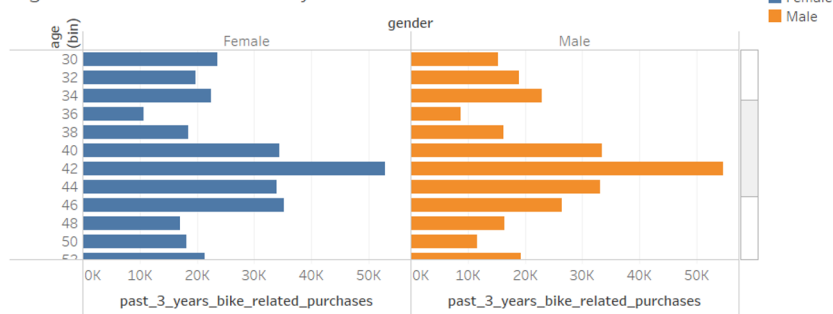
Sum of Past 3 Years Bike Related Purchases for each Age (bin). Color shows details about Gender. The view is filtered on Exclusions (Age (bin),Gender) and Gender. The Exclusions (Age (bin),Gender) filter keeps 22 members. The Gender filter excludes U.

Data Exploration

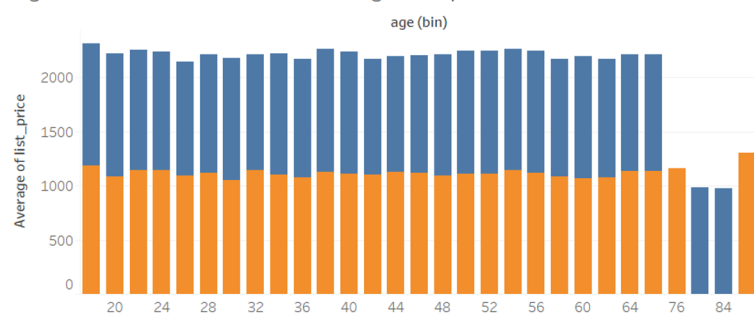
Purchase behavior by Age & Gender

- Ages bucket between **40-46** has the most intensity to buy the product.
- Age and Gender have **little impact** on change of list price

Age&Gender distribution by Purchase



Age&Gender distribution for average list_price

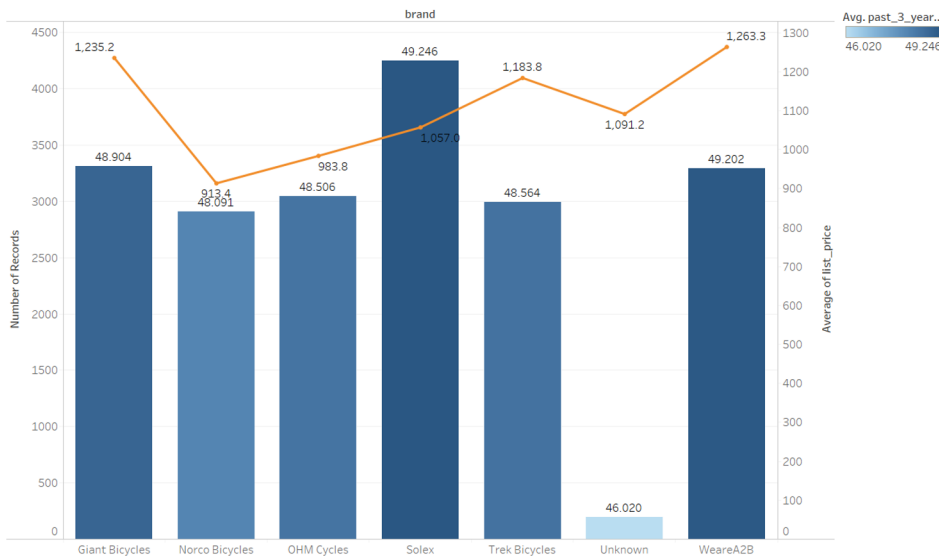


Data Exploration

Purchase Behavior by Brand

- **Solex** is the **most popular** brand in the past 3 years in terms of purchase volume (**49/yr**)
- Average purchase price of **Giant** and **WeareA2B** is ranked as NO.1 and NO.2; **higher than Solex**

Purchase Frequency and average price for each brand



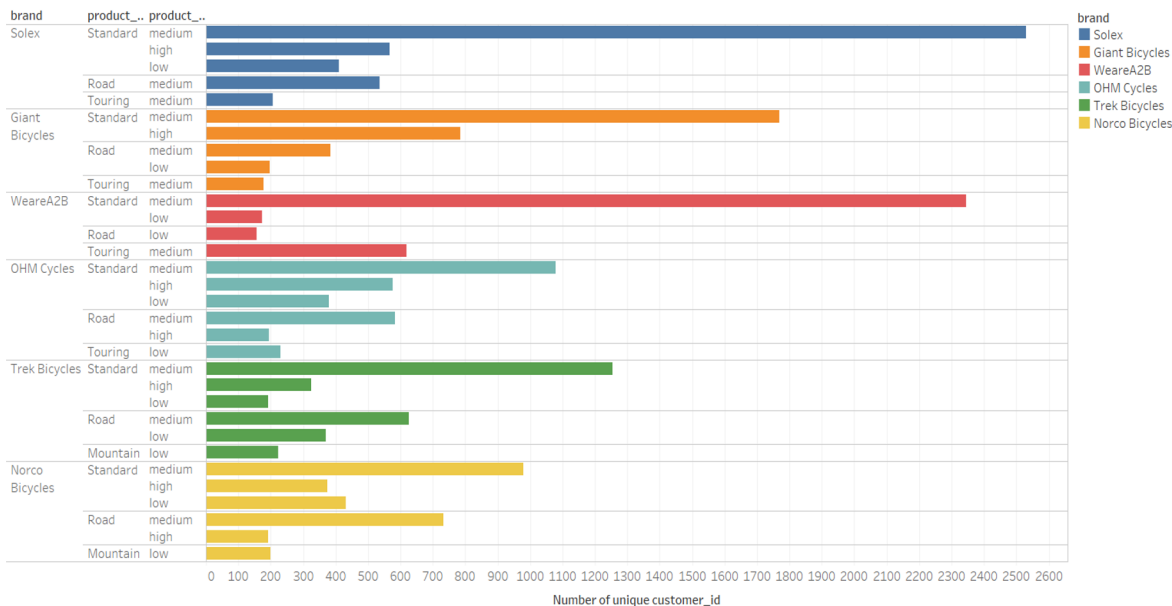
The trends of sum of Number of Records and average of list_price for brand. For pane Sum of Number of Records: Color shows average of past_3_years_bike_related_purchases. The marks are labeled by average of past_3_years_bike_related_purchases. For pane Average of list_price: The marks are labeled by average of list_price. The data is filtered on Exclusions (brand,gender), which keeps 22 members. The view is filtered on brand, which excludes Null.

Data Exploration

Top Selling Products (Class & Line)

- Overall, **Standard class medium** line product has the most selling record

Top selling product lines & classes



Count of customer_id for each product_class broken down by brand and product_line. Color shows details about brand. The view is filtered on brand, product_class and product_line. The brand filter excludes no members. The product_class filter keeps high, low and medium. The product_line filter keeps Mountain, Road, Standard and Touring.

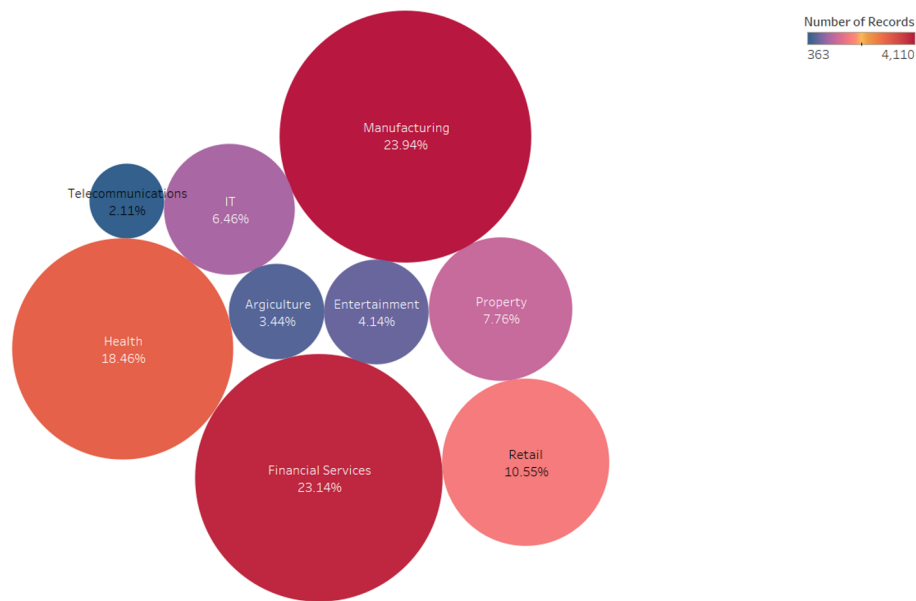
Data Exploration

Purchase Behavior in different Job Industries

Customers' job industries distribution

Top 3 Industry(Prime Customer group)

- Manufacturing(**23.94%**),
- Financial Services(**23.14%**),
- Health(**18.46%**).
- Occupy **65.54% in total**

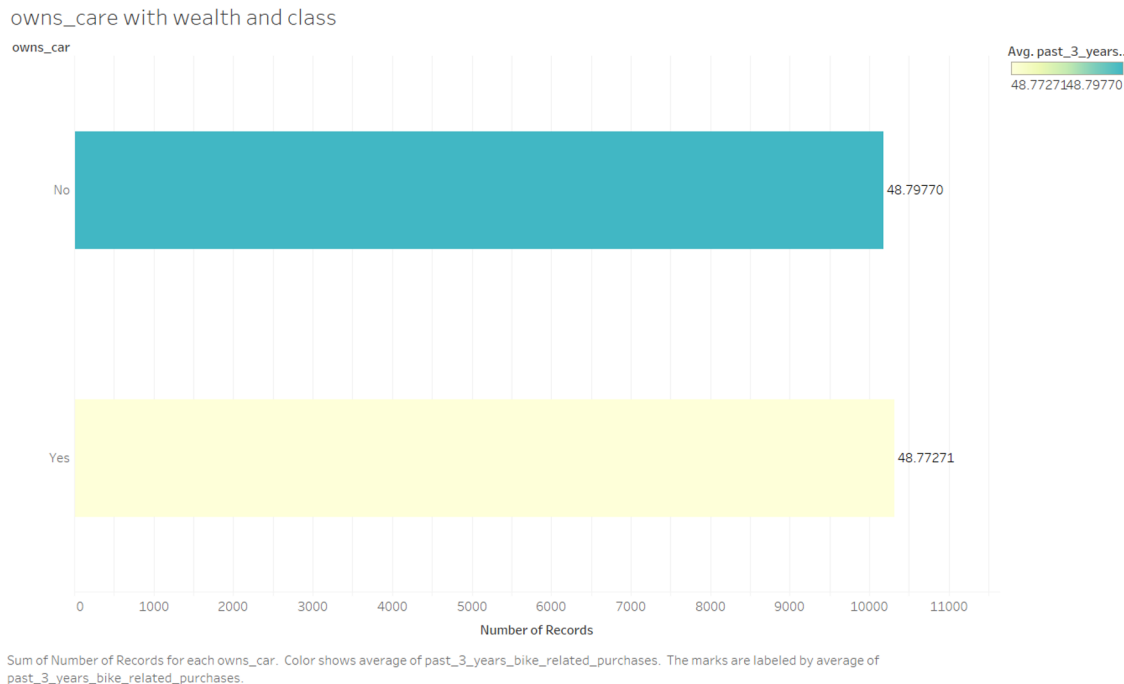


Job_industry_category and % of Total Number of Records. Color shows sum of Number of Records. Size shows % of Total Number of Records. The marks are labeled by job_industry_category and % of Total Number of Records. The view is filtered on job_industry_category, which excludes Unknown.

Data Exploration

Whether owning a car impacts the purchase behavior.

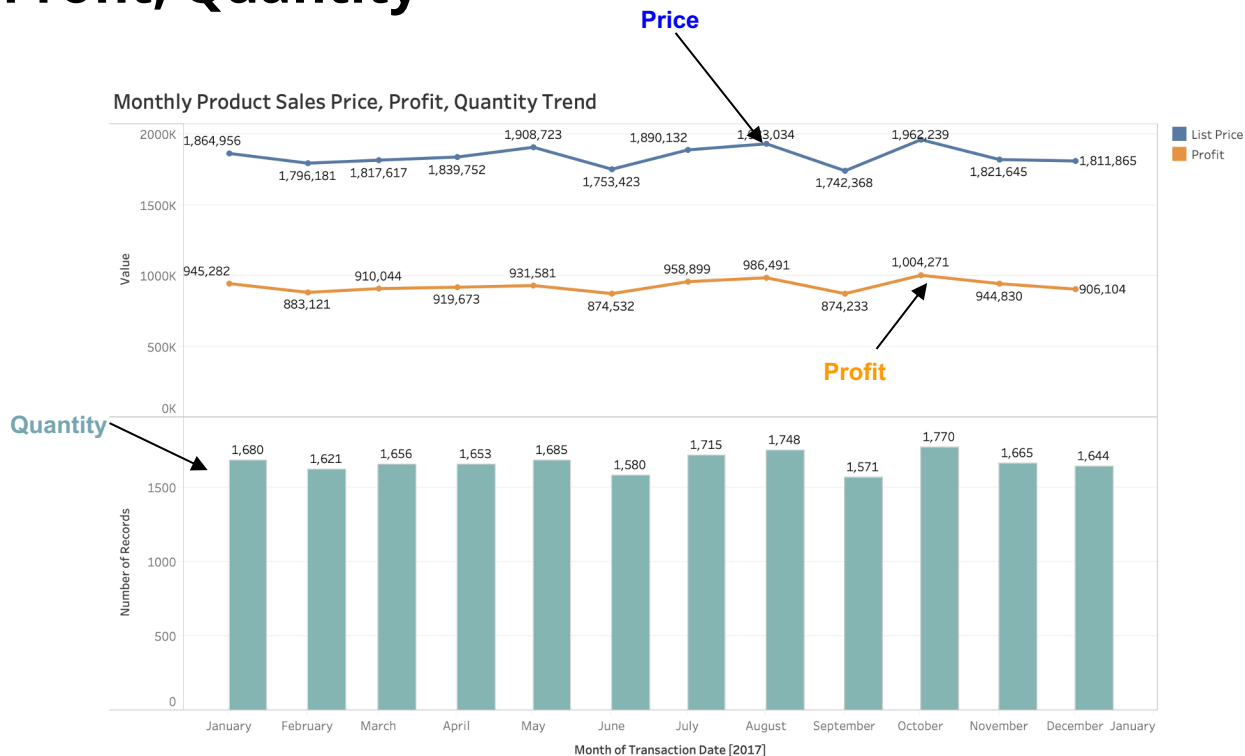
- Owning cars has **little impact** on purchasing frequency



Data Exploration

Monthly Sales Price, Profit, Quantity

- July, August, and October are 3 peaks among the year, rest of month don't show any significant fluctuations



Model Development

Segmentation Results

- After analyzing the data by different segmentation (age, job industry, product, etc),
- The total number of our targeted **HIGH-VALUE** customers is **72** from 1000
- For validation purposes, the hypothesis testing is carried in the next step

	Full Name	gender	past_3_years_bike_rel	age	DOB	job_title	job_ind	wealth	deceased	owns_c	tenure	address	postcode	state	country	property_val	Rank	Value
1	Lucine Stutt	Female		64	41-1979-01-28	Account R	Manufact	Affluent C	N	Yes	5	207 Annan	4814	QLD	Australia	1	4	1.70
2	Duff Karłowicz	Male		50	48-1972-04-28	Speech Pa	Manufact	Mass Cust	N	Yes	5	717 West l	2200	NSW	Australia	10	8	1.66
3	Winnifred Beswetherick	Female		83	44-1976-06-08	Actuary	Financial S	Mass Cust	N	No	14	61 4th Stre	3040	VIC	Australia	10	17	1.55
4	Sharron Claibourn	Female		62	40-1980-01-26	Unknown	Financial S	High Net V	N	Yes	17	555 Hermi	2280	NSW	Australia	8	30	1.46
5	Arleen Casbolt	Female		48	45-1975-05-10	Senior Que	Financial S	Affluent C	N	Yes	13	41042 Lotf	4817	QLD	Australia	1	36	1.45
6	Mitchell MacCague	Male		58	41-1979-04-11	Unknown	Manufact	Mass Cust	N	No	15	240 Acker	3190	VIC	Australia	8	38	1.44
7	Gale Disbrow	Female		59	43-1977-05-14	Cost Accoi	Financial S	Mass Cust	N	Yes	17	169 Bashfk	3741	VIC	Australia	7	46	1.41
8	Heinrick Shilstone	Male		60	42-1978-02-11	Unknown	Manufact	Affluent C	N	No	10	998 Gale P	3174	VIC	Australia	8	50	1.41
9	Hasheem Groucock	Male		98	46-1974-09-17	Budget/Ac	Manufact	High Net V	N	Yes	10	12351 Spe	2034	NSW	Australia	12	72	1.35
10	Gardie Crellim	Male		58	45-1975-09-12	Biostatistit	Manufact	High Net V	N	No	8	564 Forest	3161	VIC	Australia	12	84	1.33
11	Dolley Starmont	Female		53	42-1977-11-08	Geologist I	Manufact	Affluent C	N	No	18	4252 Dove	4129	QLD	Australia	6	114	1.28
12	Tessa Frieze	Female		89	43-1976-10-24	Health Coi	Financial S	Mass Cust	N	No	21	98158 Alpi	4212	QLD	Australia	9	157	1.20
13	Robert Corkill	Male		55	44-1976-03-12	Clinical Spi	Health	Mass Cust	N	No	4	5612 Toba	2227	NSW	Australia	10	166	1.18
14	Elvira Kurten	Female		65	47-1973-03-18	Assistant F	Financial S	High Net V	N	No	12	65 Ridge C	2471	NSW	Australia	3	166	1.18
15	Kerriann Newham	Female		71	45-1975-03-26	Assistant B	Financial S	Mass Cust	N	No	5	0193 Nort	4179	QLD	Australia	9	177	1.16
16	Dorian Rustman	Male		78	45-1974-12-09	Junior Exe	Manufact	Mass Cust	N	No	18	6156 Sumr	4352	QLD	Australia	7	181	1.16
17	Dorian Stollen	Male		78	40-1980-02-16	Statistician	Financial S	Mass Cust	N	Yes	18	72922 Can	2026	NSW	Australia	11	219	1.13
18	Kort Disley	Male		66	41-1979-02-05	Technical I	Health	Mass Cust	N	Yes	7	76 Donald	2160	NSW	Australia	9	219	1.13
19	Gothart Artus	Male		52	42-1978-05-31	Health Coi	Health	Mass Cust	N	Yes	14	21824 Nor	2117	NSW	Australia	10	231	1.11
20	Ellsworth Andrieux	Male		49	49-1971-04-26	Senior Cos	Financial S	Affluent C	N	Yes	10	08912 Car	4036	QLD	Australia	7	248	1.08
21	Sunny Christescu	Female		90	45-1975-03-12	Cost Accoi	Financial S	Mass Cust	N	No	11	6668 Blue	2209	NSW	Australia	10	250	1.08
22	Antonietta Egle	Female		82	47-1973-09-25	Actuary	Financial S	Mass Cust	N	No	7	590 Hagan	3805	VIC	Australia	7	259	1.06
23	Lark Gonet	Female		89	48-1972-01-17	Database I	Health	High Net V	N	No	8	261 Orin C	2763	NSW	Australia	8	259	1.06
24	Theresa Cowper	Female		99	44-1976-08-24	Accountan	Manufact	Mass Cust	N	No	3	88 Mifflin	2529	NSW	Australia	10	271	1.06
25	Raye Roo	Female		64	44-1976-03-07	Database I	Financial S	Mass Cust	N	Yes	16	1199 Expr	3046	VIC	Australia	9	278	1.05
26	Brynna Tivers	Female		81	46-1974-08-09	Recruiting	Financial S	Mass Cust	N	No	6	0 Mayfield	4272	QLD	Australia	7	304	1.03
27	Roch Symson	Female		96	42-1978-05-13	Office Assi	Manufact	High Net V	N	No	18	016 Westp	3073	VIC	Australia	9	312	1.02
28	Otis	Male		59	49-1971-01-11	Electrical E	Manufact	Affluent C	N	No	12	04 Oakridge	2075	NSW	Australia	11	358	0.98

Interpretation

Hypothesis Testing (Z-test)

Group 1: The AVG purchase frequency of potential High-value customers

Group 2: The AVG purchase frequency of non High-value customers (sampling out 72 record from 1000)

Null hypothesis: There is no difference between Group 1 and Group 2

Conclusion: Reject null hypothesis, **there is actually a significant difference** among two groups

z-Test: Two Sample for Means		
	Variable 1	Variable 2
Mean	72.91666667	44.81944
Known Variance	232.1338	707.812
Observations	72	72
Hypothesized Mean Dif	0	
z	7.776393503	
P(Z<=z) one-tail	3.77476E-15	
z Critical one-tail	1.644853627	
P(Z<=z) two-tail	7.54952E-15	
z Critical two-tail	1.959963985	

Defensibility

1. Could generate new features by incorporating external data like transportation, person workplace address as a stronger indicator in determining whether a particular customer would make a purchase
1. May need customers future transactional data to validate our model