MY PORTFOLIO

Khairul Ikhwan

Data Analytics | Business Intelligence

-01-

ABOUT ME

-02-

EDUCATION

-03-

PROFESSIONAL EXPERIENCE

-04-

MY PROJECT

ABOUT ME

As a professional Data Analytics who uses data to inform business decisions. This can include tasks such as collecting and analyzing data, creating reports and visualizations, and identifying patterns and trends in the data.

Completed a 3-Months Full Stack Data Analytics Certificated program from RevoU School who skilled in **SQL** (Bigquery); **Python** in Collab (Numpy, Pandas, Matplotlin, Sklearn); **Looker Studio and Tableau** (for Data Visualization Tools); **Excel** and **Spreadsheet** (for Statistical Software); Exploratory Data Analysis (**EDA**), Descriptive Statistic, Correlation Analysis, Cluster and Regression.

**Universitas Gadjah Mada
(2015-2019)**

Bachelor Degree of Tourism Department

- Thesis: Level of Customer Satisfaction using Customer Satisfaction Index (CSI) and Importance Performance Analysis (IPA) Methods
- KKN-PPM UGM: "Potential Empowerment in Galo-Galo Village become Selling Value of Tourism and Maritime Affairs.

**RevoU
(2022)**

Full Stack Data Analytics Graduate From RevoU

- Learning about Understanding business problems, Statistics and Data Exploration, data manipulation, programming language (SQL and Python), data visualization (Tableau and GDS), and Data Communication
- Taught by data analytics experts from top companies in SEA (e.g GoTo, Tokopedia, Lazada, Shopee)
- Weekly assignments include statistics and data exploration in Google Spreadsheet, SQL querying in Google BigQuery, data analytics in Google Colab using Python, and data visualization (dashboards) in Tableau and GDS.

2019

Traveloka

Flight International Ticketing, Team Leader *(Dec 2019 - Mar 2021)*

- Leading the team 42 members efficiently and effectively to meet targets with service level 85%
- Delivered report of ticketing and GDS performance by weekly
- Ensure the new processes or procedures being implemented by doing daily briefing
- Actively doing weekly 1 on 1 with the team member

2021

Vads Indonesia

Learning and Development, Trainer *(Apr 2021 - Dec 2022)*

- Designing and developing learning and development programs and materials
- Delivered training and development programs in-class or online with a passing rate of 99% of all training participants
- Build the scorecard to evaluating the effectiveness training programs
- Providing feedback by google form and coaching to employees on learning and development topics
- Maintain the tracker training records and reporting on the progress of employees by weekly
- Aligning new information into training programs by doing weekly calibration
- Create Monthly Quiz every month and analysis the errors that occurred

Project 1: EDA & Statistical Measurement

Goals

Create statistical measurement to give some insight and recommendation:

a. Aspect impact makes high impact on the pricing in Mont Kiara, Kuala Lumpur?

b. What price recommendation for 3 rooms, 2 bathrooms, 2 car parks and 100sq ft on Mont Kiara, Kuala Lumpur

Action

Clean up the Data:
   a. Remove blank column
   b. Outlier, duplicate, trim whitespace
   c. Data Cleaning
Create EDA & Statistical Measurement:
   a. The characteristic of the Property Listing
Create Statistical Measurement
   a. Using Correlation and Hypothesis testing
   b. Linear Regression (Multicolinearity)
   c. Provide price recommendation using coefficients variable

EDA & Statistical Measurement

Tools

Google Sheets

---

a. What aspect makes a high impact on the pricing in **Mont Kiara?**

CORRELATION

	Price	Rooms	Bathrooms	Car Parks	Size 2
Price	1				
Rooms	0.76	1			
Bathrooms	0.73	0.85	1		
Car Parks	0.45	0.57	0.56	1	
Size 2	0.88	0.81	0.81	0.52	1

Insights:

Based on the correlation matrix using all numerical columns, there is a positive correlation between variable. **Size is the aspect that giving the highest impact on the pricing in Mont Kiara (Strong Positive Correlation).**

Classic Assumption Test (2)

MULTICOLINEARITY

	Price	Rooms	Bathrooms	Car Parks	Size 2
Price	1				
Rooms	0.76	1			
Bathrooms	0.73	0.85	1		
Car Parks	0.45	0.57	0.56	1	
Size 2	0.88	0.81	0.81	0.52	1

But as per the problem need all independent variables to predict the dependent variable, so we still use all of the variables.

Insights:

As we see on the left, Rooms and Bathrooms have a high correlation value. **This is an indication of multicollinearity in the regression model.**

We can exclude one of the variable before model the data, or we can still use all of the variable first and eliminate the variable one by one started from the most non-significant coefficient result.

---

b. What price we can recommend for 3 rooms, 2 bathrooms, 2 car parks, and 100 sq. ft. for Properties in Mont Kiara?

SUMMARY OUTPUT

Regression Statistics	
Multiple R	0.9769916381
R Square	0.9545126609
Adjusted R Square	0.9542352991
Standard Error	489508.8972
Observations	660

	Coefficients	P-value
Intercept	0	#N/A
Rooms	99237.99479	0.000557334455
Bathrooms	-16622.85598	0.5398912822
Car Parks	-84253.327	0.006459741086
Size 2	773.4766772	0

Linear Regression Equation:

price= 99237.99 (rooms) - 16622.86 (bathrooms) - 84253.33 (car parks) + 773.48 (size)

Insights:

- Bathroom is not significantly impacted the price because the p-value > alpha-threshold (0.05)
- **Based on R-sq value (0.95), whose value close to one;** that means the independent variables **can explain (or predict) 95% of dependent variable.** High R-sq value also describes how strength the relationship between linear model and the dependent variables.
- **Recommended price** for 3 rooms, 2 bathrooms, 2 car parks, and 100 size (sq. ft.) for property in Mont Kiara is RM 173,309.29

Result

Size is the aspect that giving the **highest impact** on the **pricing** in Mont Kiara (Strong Positive Correlation).

As we see on the left, Rooms and Bathrooms have a high correlation value. This is an indication of multicollinearity in the regression model.

Recommended price for 3 rooms, 2 bathrooms, 2 car parks, and 100 size (sq. ft.) for property in Mont Kiara is **RM 173,309.29**

Project 2: Query Data using SQL BigQuery

Goals

Create Query using BigQuery SQL:

- Data Cleaning
- Aggregation Function
- Date Function
- Join
- Union

Action

Select the required data from Database. Create syntax to analyze data.

a. Get the number of unique users, number of orders, and total sale price per status and month.
b. Get frequencies, average order value and total number of unique users where status is complete grouped by month
c. Find the user id, email, first and last name of users whose status is refunded on Aug 22
d. Get the top 5 least and most profitable product over all time
e. Get Month to Date of total profit in each product categories of past 3 months (current date 15 Aug 2022), breakdown by date

Syntax

Table

Query using SQL

Tools

Google Big Query

SQL

```sql
SELECT DATE_TRUNC(DATE(created_at), MONTH) AS month_year,
       status AS status,
       COUNT(DISTINCT user_id) AS total_unique_users,
       COUNT(DISTINCT order_id) AS total_orders,
       SUM(sale_price) AS total_sales_price
FROM `bigquery-public-data.thelook_ecommerce.order_items`
WHERE created_at BETWEEN '2019-01-01' AND '2022-09-01'
GROUP BY 1,2
ORDER BY 2,1
```

JOB INFORMATION	RESULTS	JSON	EXECUTION DETAILS	
Row	month_year	status	total_unique_	total_orders	total_sales_
1	2019-01-01	Cancelled	5	5	364.919994...
2	2019-02-01	Cancelled	12	12	1072.11000...
3	2019-03-01	Cancelled	23	23	1919.62001...
4	2019-04-01	Cancelled	31	31	2629.25000...
5	2019-05-01	Cancelled	34	34	2569.82999...

```sql
SELECT DATE_TRUNC(DATE(shipped_at), MONTH) AS month_year,
       ROUND(COUNT(DISTINCT order_id)/COUNT(DISTINCT user_id),2) AS
frequency,
       ROUND(SUM(sale_price)/COUNT(DISTINCT order_id),2) AS aov,
       COUNT(DISTINCT(user_id)) AS number_of_unique_users
FROM `bigquery-public-data.thelook_ecommerce.order_items`
WHERE shipped_at BETWEEN '2019-01-01' AND '2022-09-01'
       AND status = 'Complete'
GROUP BY 1
ORDER BY 1
```

JOB INFORMATION	RESULTS	JSON	EXECUTION DETAILS	
Row	month_year	frequency	aov	number_of_
1	2019-01-01	1.0	39.07	6
2	2019-02-01	1.0	82.93	12
3	2019-03-01	1.0	67.11	28
4	2019-04-01	1.0	74.88	55
5	2019-05-01	1.01	69.2	72

```sql
SELECT DISTINCT (u.id) AS user_id,
       u.email AS email,
       u.first_name,
       u.last_name
FROM `bigquery-public-data.thelook_ecommerce.orders` AS o
LEFT JOIN `bigquery-public-data.thelook_ecommerce.users` AS u
ON o.user_id = u.id
WHERE o.returned_at BETWEEN '2022-08-01' AND '2022-09-01'
       AND o.status = 'Returned'
ORDER BY 3
```

JOB INFORMATION	RESULTS	JSON	EXECUTION DETAILS	
Row	user_id	email	first_name	last_name
1	95357	aaroncastro@example.com	Aaron	Castro
2	66413	abigaillopez@example.com	Abigail	Lopez
3	32721	adammiller@example.net	Adam	Miller
4	36226	adamowens@example.com	Adam	Owens
5	59653	adamrobinson@example.com	Adam	Robinson

Query using SQL

Tools

Google Big Query

SQL

```sql
WITH product_sales AS
( SELECT i.product_id,
         p.name,
         p.retail_price,
         p.cost,
         i.sale_price - p.cost as profit
  FROM `bigquery-public-data.thelook_ecommerce.order_items` i
  JOIN `bigquery-public-data.thelook_ecommerce.products` p
  ON i.product_id = p.id
  WHERE status = 'Complete'),

product_profit AS
( SELECT product_id,
         name,
         retail_price,
         cost,
         SUM(profit) AS total_profit
  FROM product_sales
  WHERE name IS NOT NULL
  GROUP BY 1,2,3,4
  ORDER BY 2)
```

Row	product_id	name	retail_price	cost	total_profit
1	14235	Indestructable Aluminum Alum...	0.01999999...	0.00829999...	0.03509999...
2	14159	Set of 2 - Replacement Insert For Checkbook Wallets Card Or Picture Insert	0.49000000...	0.17738000...	0.31262000...
3	12536	Individual Bra Extenders	1.75	1.01324999...	0.73675000...
4	13629	Solid Color Leather Adjustable ...	1.50999999...	0.64476999...	0.86522999...
5		Black Lace Cuff Beanie Cap (C...	1.82000005...	0.89180002...	0.92820002...

```sql
(SELECT *
FROM product_profit
ORDER BY 5 DESC
LIMIT 5)
UNION ALL
(SELECT *
FROM product_profit
ORDER BY 5 ASC
LIMIT 5)
```

```sql
WITH profit AS
( SELECT DATE(o.shipped_at) AS order_date,
         p.category AS product_category,
         ROUND(SUM(o.sale_price - p.cost),2) AS category_profit
  FROM `bigquery-public-data.thelook_ecommerce.order_items` AS o
  INNER JOIN `bigquery-public-data.thelook_ecommerce.products` AS p
  ON o.product_id = p.id
  WHERE o.shipped_at BETWEEN "2022-01-01" AND "2022-08-16" AND status =
"Complete"
  GROUP BY 1,2
  ORDER BY 2,1
),

mtd_table AS
( SELECT order_date,
         product_category,
         category_profit,
         SUM(category_profit) OVER(PARTITION BY product_category, EXTRACT(MONTH
FROM order_date) ORDER BY product_category,order_date) AS mtd
  FROM profit
  ORDER BY 2,1
```

JOB INFORMATION		RESULTS	JSON	EXECUTION DETAILS

Row	order_date	product_category	mtd
1	2022-06-15	Accessories	1174.21999...
2	2022-07-15	Accessories	2003.47999...
3	2022-06-15	Active	1076.11
4	2022-07-15	Active	2392.62999...
5	2022-08-15	Active	2223.43

```sql
SELECT order_date,product_category, mtd
FROM mtd_table
WHERE order_date BETWEEN "2022-06-01" AND "2022-08-16"
      AND EXTRACT(DAY FROM order_date) = 15
```

Project 3: Programming Data

Goals

Suggest Marketing Campaign from Business Analyst:

- The marketing and sales team would like to run a thematic communication campaign
- Clustering a few different relevant segmentation
- Provide recommendations for the themes of the campaign.

Action

Select the required data from Database. Create syntax to analyze data.

Data Preparation
a. Import Libraries and Dataset
b. Checking missing values
c. Handling missing values
d. Converting data type
e. Removing outlier
f. Combine Dataset

Exploratory Data Analyst (EDA)
a. Create aggregation each variable
b. Explain the data with Grafik and Numerik
c. Highlight trend and poin anomali data

Clustering and Insight Recommendation
a. Check The Distributions Data
b. Elbow Method
c. Silhouette Method
d. Cluster and characteristic result
e. Recommendation

Programming Data

Result

Tools

- Spreadsheet for dataset
- Google Collab Python

Recommendation:
- There are **4 segments** formed based on the socio-economic background of the user and focus to cluster **Loyal**.
- **Government bond investment** have the **same product** character as the **bond mutual fund**. Therefore, to expand into government bond investment products, we can create a campaign to **Loyal Customer** due transaction on bond mutual fund the **highest investment**.
- Long Invest for Long Life: Provide benefits for **long-term investment**.
- Mutual Points: Provide **cashback promos in non-cash form**, which can be in the form of points that can only be used for transactions on our platform. So that it can **encourage** them to continue making transactions on our platform.

Project 4: Interactive Dashboard Data Visualization

Goals

Create interactive report and interactive dashboard AirBnB Listings in Singapore.

- Create scorecard number
- There are filter by date, location, aggregation listing
- Create grafik each variable that can show all the data

Action

Import Dataset from database excel, spreadsheet or Query SQL
- Combine dataset that have same primary key
- Aggregation function

Create new sheet
- Exploratory Data Analyst (EDA)
- Create aggregation each variable
- Explain the data with Grafik and Numerik
- Highlight trend and poin anomali data

Summary Dashboard
- Combine element all sheet data visualization
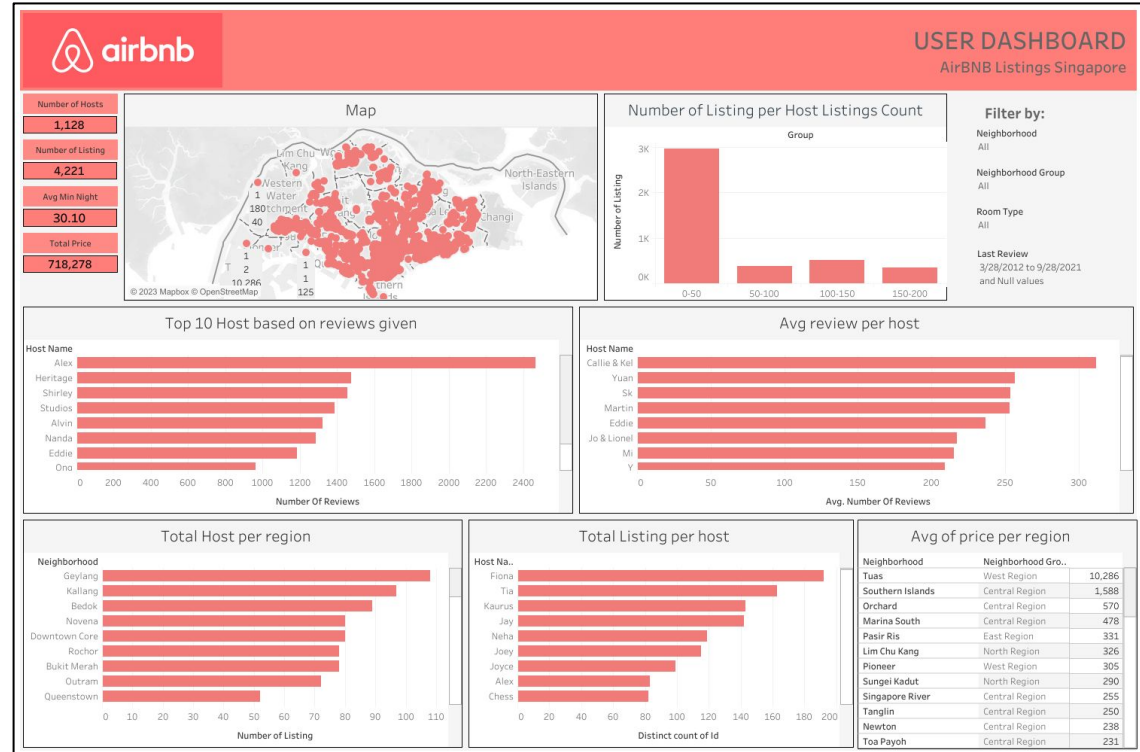- Create interactive dashboard

CONTACT

khairul.ikhwaan@gmail.com

+62 852 6081 7579

https://www.linkedin.com/in/khairul-ikhwan-/

@khairulikwann