

Holiday Package

Prediction

Yuk Liburan! (Data Scientist)

Final Presentation Kelompok 5 - Batch 18



Meet our team



- Aisyah Khairunnisa A
- Budi Dwi Ananto
- David Prayogo
- Gayatri A
- Hafidh Rizky
- Muharlan
- Shafa Amelia
- Sony Monthona R K



Outline



- Business Understanding
- Data Overview, Exploratory Data Analysis & Insight
- Data Cleaning & Preparation
- Model & Evaluation
- Business Insight & Recommendation



Business Understanding

Problem Statement



4888

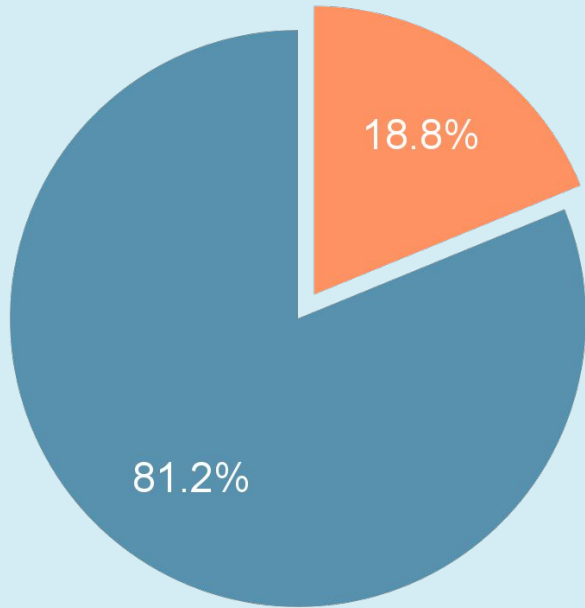
Jumlah semua pelanggan

920

Jumlah pelanggan yang
membeli paket liburan

Perusahaan Trips&Travel.com ingin meluncurkan paket liburan baru yang bernama Wellness Tourism Package. Sayangnya mereka memiliki masalah dengan promosi. Mereka menghubungi pelanggan secara acak untuk menawarkan paket liburan sehingga tidak efisien.

Let's check the data!



- Membeli paket liburan
- Tidak membeli paket liburan

- Berdasarkan data tahun lalu, Trips&Travel.com memiliki pelanggan sebanyak 4888 orang.
- Namun, hanya 18% dari total pelanggan yang membeli paket liburan.

Goal, Objective, Business Metrics



Goal:

Melakukan promosi secara efisien sehingga dapat mengoptimalkan penggunaan biaya.

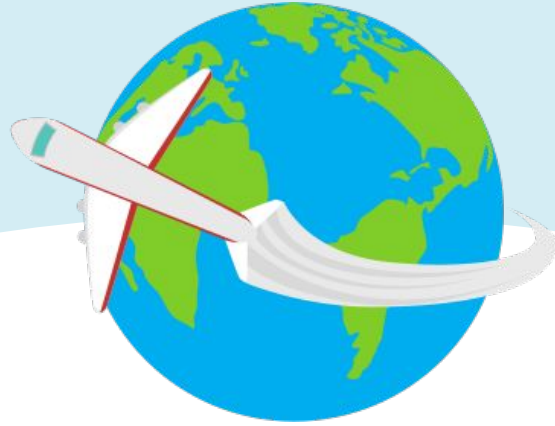
Objective:

Membuat model machine learning untuk memprediksi user yang berpotensi untuk membeli paket liburan.

Business Metrics:

- Persentase user yang membeli paket liburan
- CAC (customer acquisition cost)
- Revenue





Data Overview, Exploratory Data Analysis & Insight

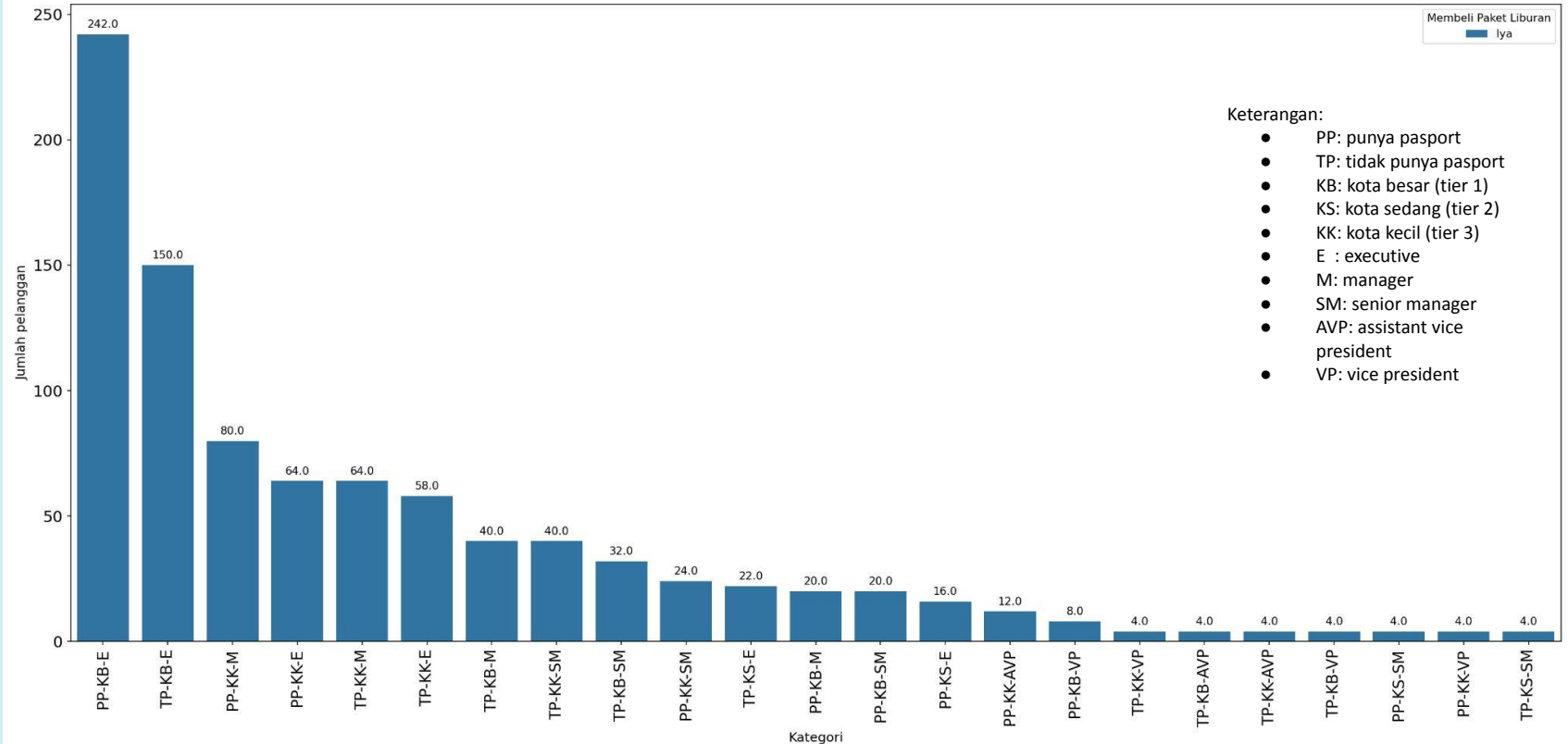
Data Overview



No.	Variabel	No.	Variabel
1.	Customer ID	11.	Product Pitched
2.	Prod Taken	12.	Preferred Property Star
3.	Age	13.	Marital Status
4.	Type of Contact	14.	Number of Trips
5.	City Tier	15.	Passport
6.	Duration of Pitch	16.	Pitch Satisfaction Score
7.	Occupation	17.	Own Car
8.	Gender	18.	Number of Children Visiting
9.	Number of Person Visiting	19.	Designation
10.	Number of Follow Ups	20.	Monthly Income

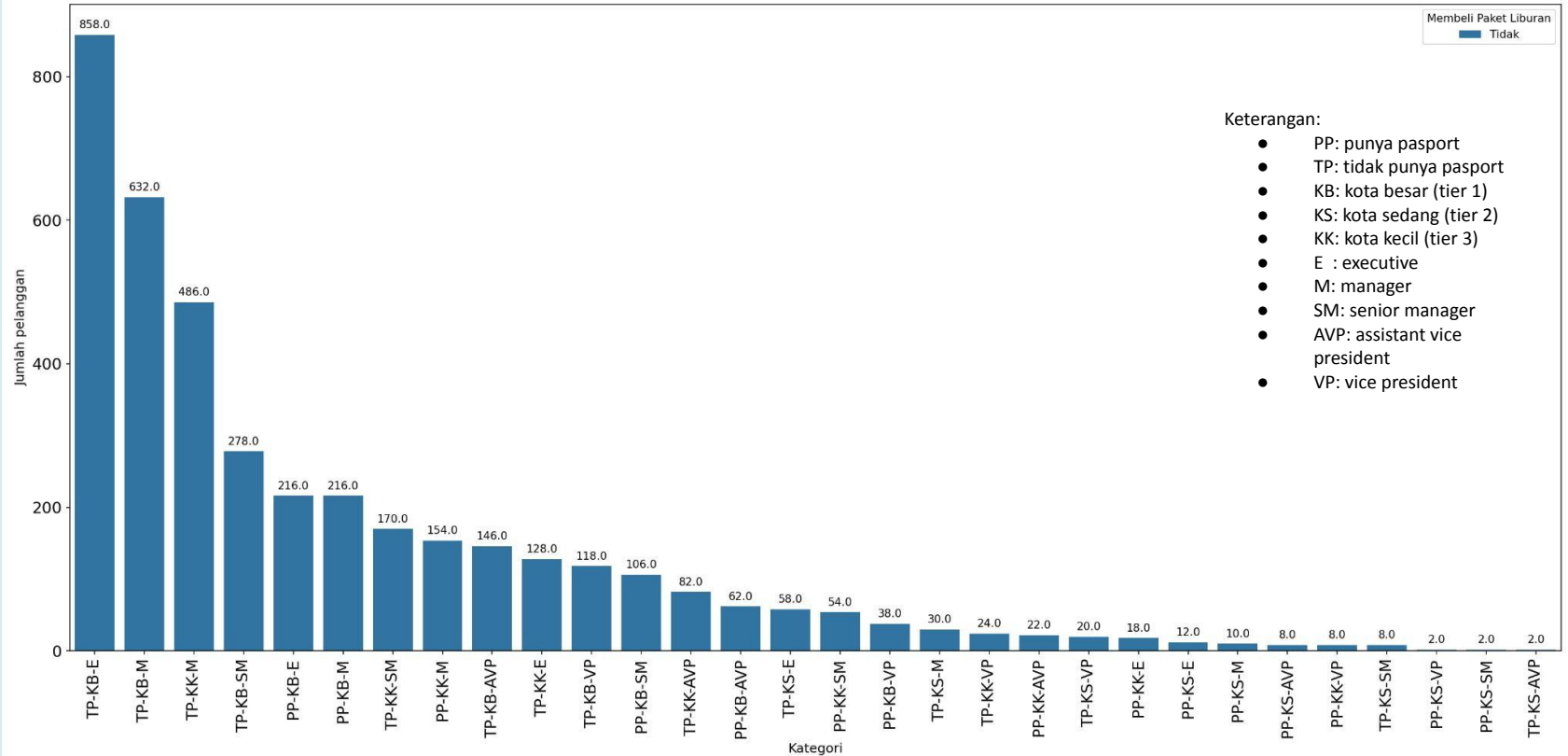
- Target : Prod Taken
- Fitur kategorikal : 6 fitur
- Fitur numerikal : 14 fitur

Exploratory Data Analysis

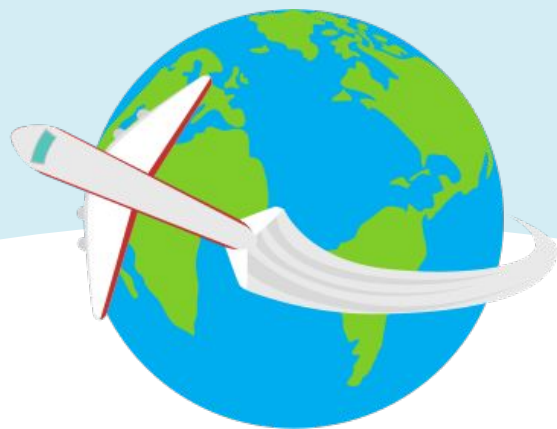


Grafik Jumlah Pelanggan yang Mengambil Paket Liburan berdasarkan Fitur Paspor, Tingkat Kota, Jabatan

Exploratory Data Analysis



Grafik Jumlah Pelanggan yang tidak Mengambil Paket Liburan berdasarkan Fitur Paspur, Tingkat Kota, Jabatan

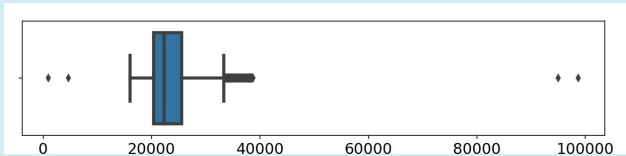


Data Cleaning & Preparation

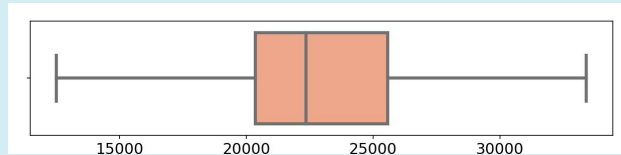
Data Cleaning



- Handling missing values & imputing → menghapus missing values <1%, 2 fitur diimputasi dengan mean, 1 fitur diimputasi dengan median, dan 3 fitur diimputasi dengan modus.
- Handling duplicated data → 140 data duplikat.
- Handling outlier (IQR method).



Sebelum Handling Outlier



Setelah Handling Outlier

**Data sebelum &
sesudah data cleaning
(baris):**

4888 → 4652

Data Pre-processing



Split data training & testing:
70:30

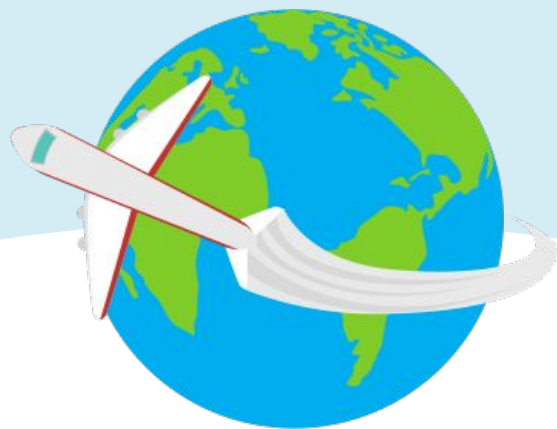
Scaling & feature transformation:
Standardization

Feature encoding:

- Label encoding
- One hot encoding

Balancing data:
SMOTE ratio 1





Model & Evaluation

Modeling



Metrics yang digunakan:

- Recall
- Precision ≥ 0.5

Karena tidak ingin nilai **false negative** yang besar (pelanggan yang sebenarnya ingin membeli paket liburan namun dianggap tidak ingin membeli).

$$\text{Recall} = \frac{\text{True positive}}{\text{True positive} + \text{False negative}}$$

$$\text{Precision} = \frac{\text{True positive}}{\text{True positive} + \text{False positive}}$$

Model After Hyperparameter Tuning



Model	Accuracy		Precision		Recall	
	Train	Test	Train	Test	Train	Test
Logistic Regression	0.74	0.76	0.74	0.41	0.75	0.80
Decision Tree	0.76	0.72	0.75	0.36	0.79	0.72
Random Forest	0.80	0.81	0.82	0.48	0.77	0.70
kNN	1.00	0.93	1.00	0.82	1.00	0.78
AdaBoost	0.88	0.85	0.90	0.58	0.85	0.55
XGBoost	0.83	0.82	0.84	0.50	0.80	0.72

Selisih tidak terlalu besar

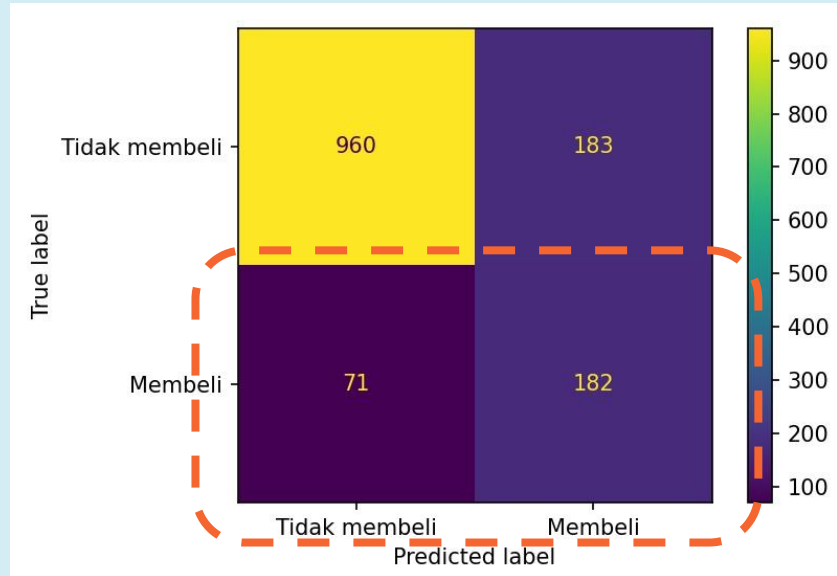
Model XGBoost dipilih karena:

- Best fit, dilihat dari score recall-nya
- Score precision-nya ≥ 0.5

Hyperparameter tuning yang dilakukan di XGBoost:

- cross validation
- max_depth
- gamma
- learning_rate
- tree_method
- colsample_bytree
- subsample
- early_stopping_rounds

Confusion Matrix

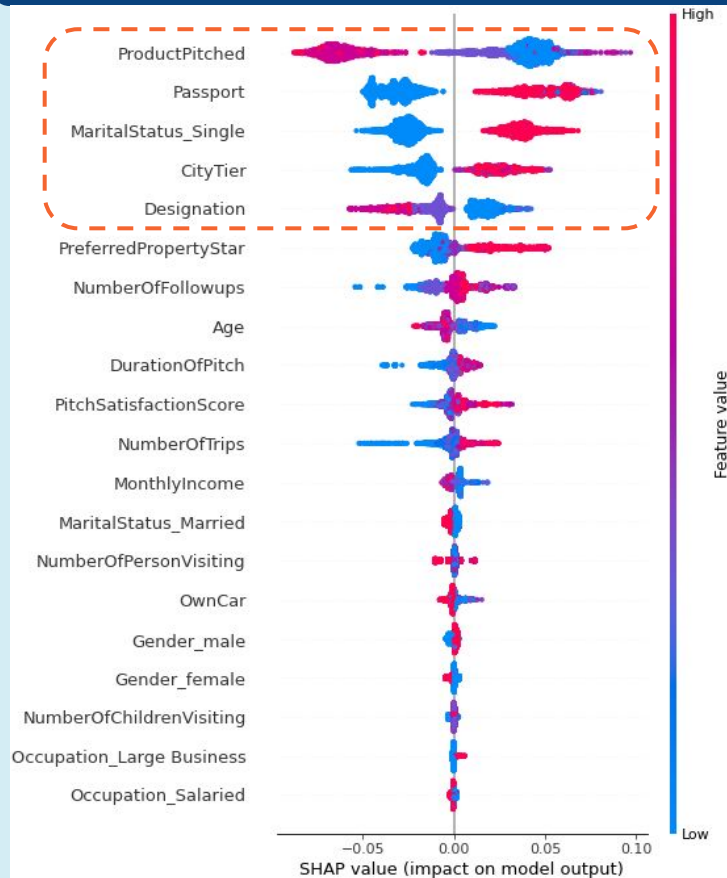


72%

Expected performance

Dari total 253 pelanggan yang membeli paket liburan, sebanyak 182 diprediksi dengan benar.

SHAP Value



Jenis produk yang diambil (product pitch):

- 0: basic
- 1: standard
- 2: deluxe
- 3: super deluxe
- 4: king

Jabatan (designation):

- 0: executive
- 1: manager
- 2: senior manager
- 3: AVP
- 4: VP

Status lajang (single):

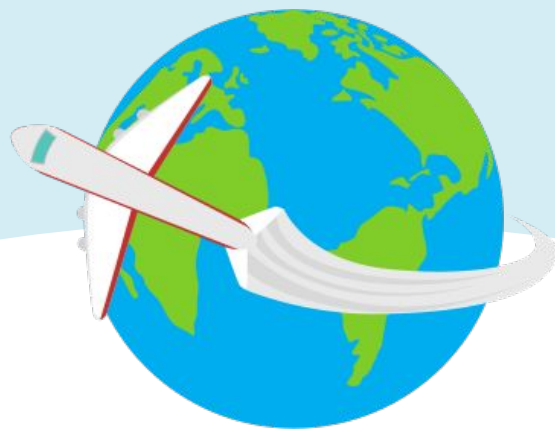
- 0: tidak lajang
- 1: lajang

Tingkat kota (city tier):

- 1: kota besar
- 2: kota sedang
- 3: kota kecil

Kepemilikan paspor (passport):

- 0: tidak punya
- 1: punya



Business Insight & Recommendation

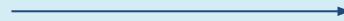
Before & After Using Model



Sebelum menggunakan model, persentase jumlah pembeli paket liburan hanya 18% namun setelah menggunakan model hasilnya meningkat jadi 72%.

Product Taken

18%



Product Taken

72%

Meningkat 4x lipat
*expected performance

Untuk melihat dampak dari model, kita menggunakan metode CAC dengan deskripsi seperti berikut.

Before Using Model



Data asli

Jenis promosi	Biaya	Jumlah target (orang)	Rincian biaya
Telepon	\$1	4.888	\$4.888
		Total biaya yang dikeluarkan	\$4.888
		Jumlah pelanggan yang membeli paket liburan	920
		CAC	\$5.31

Dari total 4.888 pelanggan yang diberi tawaran, hanya 920 pelanggan yang membeli paket liburan (18%).

**Biaya yang dikeluarkan/orang \$1*

CAC (customer acquisition cost)

After Using Model



Jenis promosi	Biaya	Jumlah target (orang)	Rincian biaya
Telepon	\$1	4.888	\$4.888
		Total biaya yang dikeluarkan	\$4.888
		Jumlah pelanggan yang membeli paket liburan	3.519
		CAC	\$1.38

Dengan asumsi model dapat memprediksi dengan tepat hingga 72% maka sebanyak 3519 pelanggan akan membeli paket liburan.

**Biaya yang dikeluarkan/orang \$1*

CAC (customer acquisition cost)

After Using Model



Total Customer Acquisition Cost			
Skenario	Jumlah pelanggan yang membeli paket liburan	CAC	Total
Before	920	\$5.31	\$4.876
After	920	\$1.38	\$1.269

Before
 $500 \times 920 - \$4.876 = \455.124



After
 $500 \times 920 - \$1.269 = \458.731

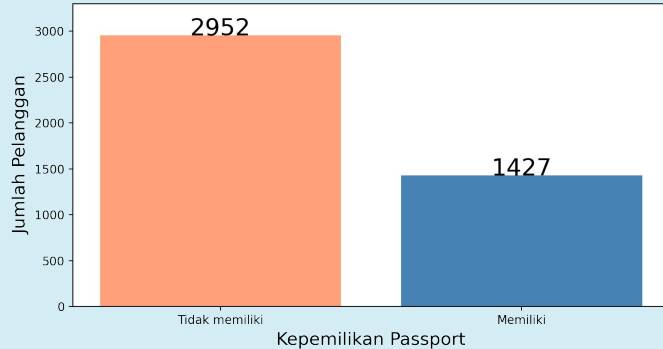
Selisih Pendapatan
 $\$458.731 - \$455.124 = \mathbf{\$3.607}$

Asumsi jika yang membeli paket berjumlah 920 orang dan harga \$500/paket Wellness

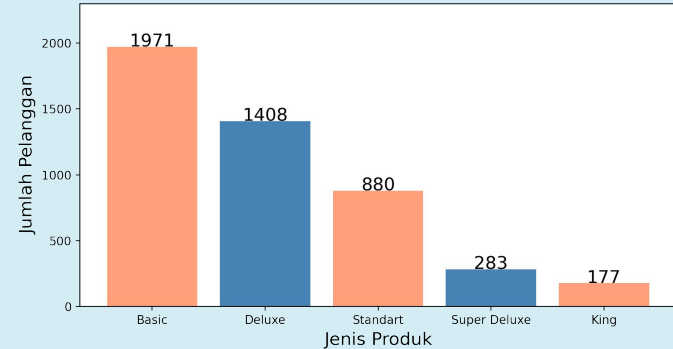
Suggestions



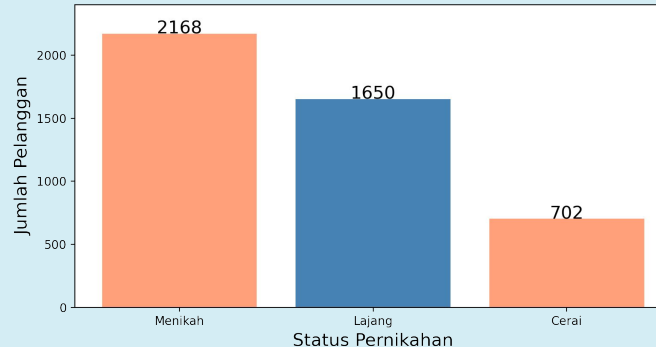
Jumlah Pelanggan berdasarkan Kepemilikan Passport pada Data Training



Jumlah Pelanggan berdasarkan Jenis Produk yang dipilih pada Data Training



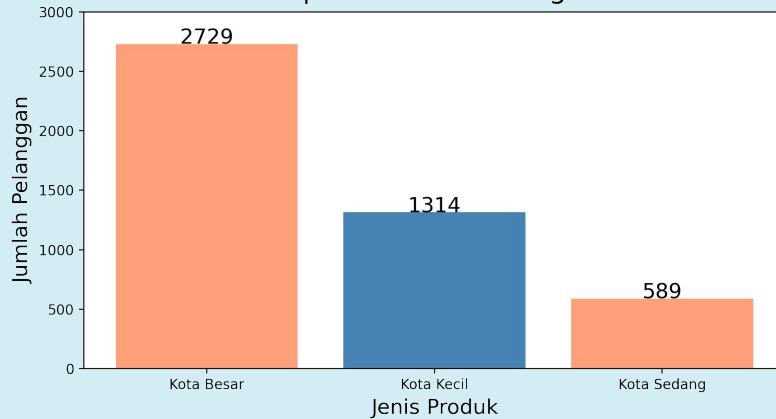
Jumlah Pelanggan berdasarkan Status Pernikahan pada Data Training



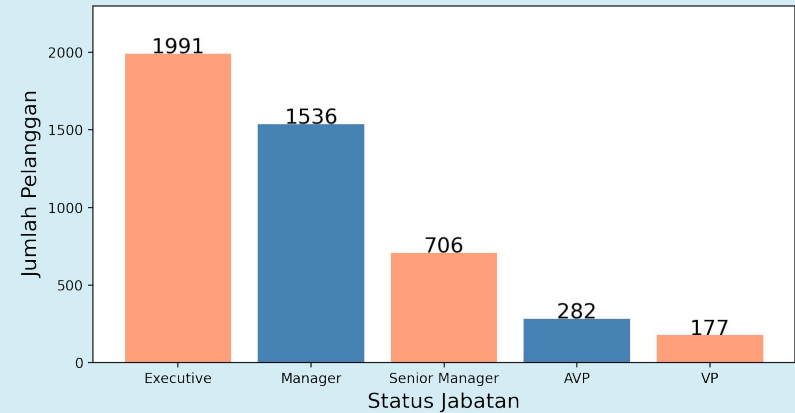
Suggestions



Jumlah Pelanggan berdasarkan Tingkat Kota pada Data Training



Jumlah Pelanggan berdasarkan Status Jabatan pada Data Training



Business Recommendations



Ciri-ciri pelanggan yang membeli paket liburan berdasarkan hasil model:

- Status lajang
- Tinggal di kota besar (tier 1) atau sedang (tier 2)
- Memiliki jabatan Executive atau Manager
- Pelanggan yang ditawarkan produk Basic & Standard
- Memiliki paspor



Action point (*High Potential Customer*)



- Menawarkan paket liburan destinasi dalam & luar negeri untuk pelanggan yang **mempunyai paspor**.
- Menawarkan paket Wellness dengan harga yang murah, seperti **Basic & Standard** kepada pelanggan yang belum pernah membeli paket liburan dan bagi yang memiliki jabatan **Executive & Manager**.
- Menggencarkan promosi kepada pelanggan yang berasal dari **kota besar & sedang**.
- **Menawarkan pelanggan yang masih lajang**, disarankan memberi promosi paket liburan di luar hari libur besar.



Action point (*Low Potential Customer*)



- Menawarkan Paket Wellness destinasi dalam negeri untuk pelanggan yang **tidak mempunyai paspor**.
- Menawarkan Paket Wellness dengan range harga menengah ke atas untuk yang memiliki posisi jabatan setidaknya **senior manager**.
- Pelanggan yang **sudah berkeluarga** mendapatkan promosi Paket Wellness di hari libur besar (seperti Hari Raya, Tahun Baru dll).



Thank You



Holiday Package

Prediction

