Methodology Document

Technical Specifications

Physical

S.No.	Machine Type	Model	Processor	RAM	GPU
2.	Google Cloud		24 Cores	60 GB	None

Software

S.No.	OS/Software	Version	Details (any specifics)	URL
1.	*Ubuntu 18.04*	*18.04 LTS*		

Data Cleaning

S.No.	Column Name	Treatment	Details
1.	season_holidayed_code, state_code_residence	Missing Values	None Used. Missing values were less and Models built were all capable of handling missing values

Feature Engineering

Transformation

Columns Name	Transformation
• All Categorical Columns	Frequency Encoding applied for LightGBM and XGBoost Label Encoding and Frequency Encoding for CatBoost

Derived Variable

Transformation Details	New Columns Name
• **checkin_date - booking_date**	1. booking_in_advance_days

● Date Variables From: **booking_date** **checkin_date** **checkout_date**	1. checkin_date_week 2. checkin_date_month checkin_date_year 3. checkin_date_dayofweek 4. checkin_date_dayofmonth 5. checkin_date_dayofyear 6. checkout_date_week 7. checkout_date_dayofweek 8. booking_date_week 9. booking_date_dayofyear
● Feature Interactions: **Categorical-Categorical**	1. resort_id_checkin_date 2. resort_id_checkout_date 3. resort_id_booking_date 4. resort_id_memberid 5. resort_id_state_code_resort 6. resort_id_state_code_residence 7. persontravellingid_member_age_buckets 8. persontravellingid_state_code_residence 9. persontravellingid_state_code_resort 10. member_age_buckets_state_code_residence 11. member_age_buckets_state_code_resort 12. state_code_residence_state_code_resort 13. memberid_channel_code 14. memberid_booking_type_code 15. memberid_reservationstatusid_code 16. memberid_resort_type_code 17. memberid_cluster_code 18. memberid_room_type_booked_code 19. memberid_checkin_date 20. memberid_checkout_date 21. memberid_booking_date 22. memberid_checkin_date_month 23. memberid_checkin_date_week 24. memberid_checkin_date_dayofweek 25. memberid_checkout_date_week 26. memberid_checkout_date_dayofweek 27. resort_id_booking_roomnights 28. memberid_booking_roomnights
● Feature Interactions: **Numerical-Numerical**	1. numberofadults_div_numberofchildren 2. numberofadults_div_total_pax 3. numberofadults_div_roomnights 4. numberofadults_div_booking_in_advance_days 5. numberofadults_div_booking_roomnights 6. numberofadults_div_total_persons_travelling 7. numberofchildren_div_total_pax 8. numberofchildren_div_roomnights 9. numberofchildren_div_booking_in_advance_days 10. numberofchildren_div_booking_roomnights 11. numberofchildren_div_total_persons_travelling

	12. total_pax_div_roomnights 13. total_pax_div_booking_in_advance_days 14. total_pax_div_booking_roomnights 15. total_pax_div_total_persons_travelling 16. roomnights_div_booking_in_advance_days 17. roomnights_div_booking_roomnights 18. roomnights_div_total_persons_travelling 19. booking_in_advance_days_div_booking_roomnights 20. booking_in_advance_days_div_total_persons_travelling 21. booking_roomnights_div_total_persons_travelling 22. roomnights_dif_booking_roomnights 23. total_persons_travelling_dif_total_pax 24. checkout_date_week_dif_checkin_date_week 25. checkout_date_week_dif_booking_date_week 26. checkin_date_week_dif_booking_date_week
● Feature Interactions: **Categorical-Numerical**	GroupBy on Category followed by Aggregation on many Numerical columns

Model Run

Run No.	Model	Metric	Value	Hyperparameter values
1	LIGHTGBM	RMSE		'n_estimators': 20000, 'learning_rate': 0.01, 'num_leaves': 48 or 100, 'colsample_bytree': 0.5000000000000001, 'subsample': 1.0, 'min_child_weight': 150, 'boosting_type': 'gbdt'
2.	XGBoost	RMSE		'learning_rate':0.01, 'n_estimators': 10000, 'max_depth': 6, 'colsample_bytree': 0.5000000000000001, 'subsample': 1.0, 'gamma' : 0.5, 'min_child_weight': 150,

3.	CatBoost	RMSE		eval_metric='RMSE', n_estimators=20000, od_type='Iter', od_wait=200, colsample_bylevel=0.7, max_depth=6, learning_rate=0.1
4.	Ensemble(Linear Regression)	RMSE	**Train: 95.29** **Test: 95.06**	default params

Coding Details

S.No	Programming Language	Package Used	Details
1.	Python	Scikit learn, Numpy, Pandas, xgbooxt, catboost, lightgbm, joblib	

Platforms/Tools Used (if any)

S.No	Platform Tool	Details