**Описание задачи**

После суровых трудовых будней всегда хочется с удовольствием провести свой отпуск и насладиться отдыхом. Тем не менее, на этом пути к счастью обычно возникают несколько преград: выбор желаемой гостиницы, заблаговременное бронирование, оплата, получение максимально возможной скидки и т.д. Как минимизировать свои риски и траты, и при этом получить желаемый уровень сервиса и отдых?

А что, если проанализировать данные по оплатам, бронированиям и отменам для сети гостиниц и попытаться найти скрытые зависимости с датами заезда, формой оплаты и другими признаками?

А если на этом не останавливаться и попытаться использовать найденные паттерны для создания нового банковского продукта, который сможет как помочь потенциальным клиентам с организацией отпуска, так и позволит банку и отелю остаться в плюсе?

Именно для создания такого инновационного продукта Вышка Онлайн совместно с Цифровой лабораторией ПСБ проводит хакатон **ВШЭ ПСБ.Хак**.

Соревнование пройдет с 20-22 сентября 2024 в гибридном формате. Командам участникам необходимо пройти через два этапа:

* *Этап 1. Технический*. Решение задачи бинарной классификации с целью предсказания факта отмены бронирования
* *Этап 2. Бизнесовый*. На основе найденных зависимостей в рамках первого этапа, а также с учетом самостоятельно проведенного анализа рынка и существующих банковских и онлайн-туристических продуктов и программ лояльности, предложить новый банковский продукт для банка **ПСБ**.

В рамках первого этапа команды подгружают свои предсказания отмены бронирования на платформу, где происходит автоматическая проверка ответов на тестовой выборке. Первый этап соревнования завершается 22 сентября в 16:00.

Детали регистрации на платформе и особенности отправки решений будут подробно описаны после торжественного открытия хакатона 20 сентября.

В рамках второго этапа команды готовят презентацию предлагаемого продукта, которую будут защищать перед жюри 22 сентября, начиная с 17:00.

Для погружения участников в особенности банковского бизнеса и предметной области в течение субботы будет проводится серия мастер-классов. Отдельно с командами будут работать менторы, которые будут консультировать на протяжении всего периода соревнования.

**Данные**

Данные представляют собой информацию о бронировании номеров различных категорий в разных гостиницах. Гостиницы с номерами 1 и 2 – один регион РФ, с номерами 3 и 4 – другой регион РФ. Структура данных следующая:

*№ брони* – идентификатор брони  
*Номеров* – количество номеров в бронировании  
*Стоимость* – стоимость номеров в рублях  
*Внесена предоплата* – сумма внесенной предоплаты  
*Способ оплаты* – один из 12 способов оплаты  
*Дата бронирования* – дата бронирования с точностью до минуты  
*Дата отмены* – дата отмены бронирования с точностью до минуты, если было  
*Заезд* – дата заезда с точностью до дня  
*Ночей* – количество ночей  
*Выезд* – дата выезда с точностью до дня  
*Источник* – онлайн-канал продаж  
*Статус брони* – один из 5 статусов  
*Категория номера* – описание категории номера. Если бронировалось несколько номеров, то идет сплошное описание с нумерацией.  
*Гостей* – число гостей  
*Гостиница* – номер гостиницы

Целевое поле – *Дата отмены*: если поле заполнено, то это соответствует целевому значению **1**, иначе – **0**. Статусы в поле *Статус брони* уточняют состояние и носят справочный характер (но могут помочь при построении модели).

В тестовой выборке отсутствуют столбцы *Дата отмены* и *Статус брони*.

**Метрика**

Итоговой метрикой выступает [ROC-AUC](https://scikit-learn.org/stable/modules/generated/sklearn.metrics.roc_auc_score.html) по столбцу *Дата отмены*.

**Лидерборд**

Лидерборд технической части соревнования (метрика модели) в процессе соревнования считается по доступной для скачивания тестовой выборке. Окончание отправок всех решений 22 сентября в 16:00. Лидерборд скрывается от участников 22 сентября в 12:00 и открывается с итоговым результатом 22 сентября в 20:00.

**Итоговая оценка решения**

Итоговая оценка соревнования формируется на основе технической части (40% результата) и защит решений в форме презентаций (60% результата). Экспертное жюри определяет оценку защит. Детали с критериями оценки второго этапа будут предоставлены командам отдельно.