Methodology:

At first glance, when we looked only at the raw averages, it seemed like resorts with snow parks had lower prices. In the control group (378 resorts) the mean ticket price was **51.39**, while in the treated group (121 resorts) it was only **40.39**. That gives us a naive ATT of about **–11**. This was a bit shocking, since the intuitive expectation is that adding extra attractions like a snow park should make a resort *more* attractive and allow it to charge *higher* prices.

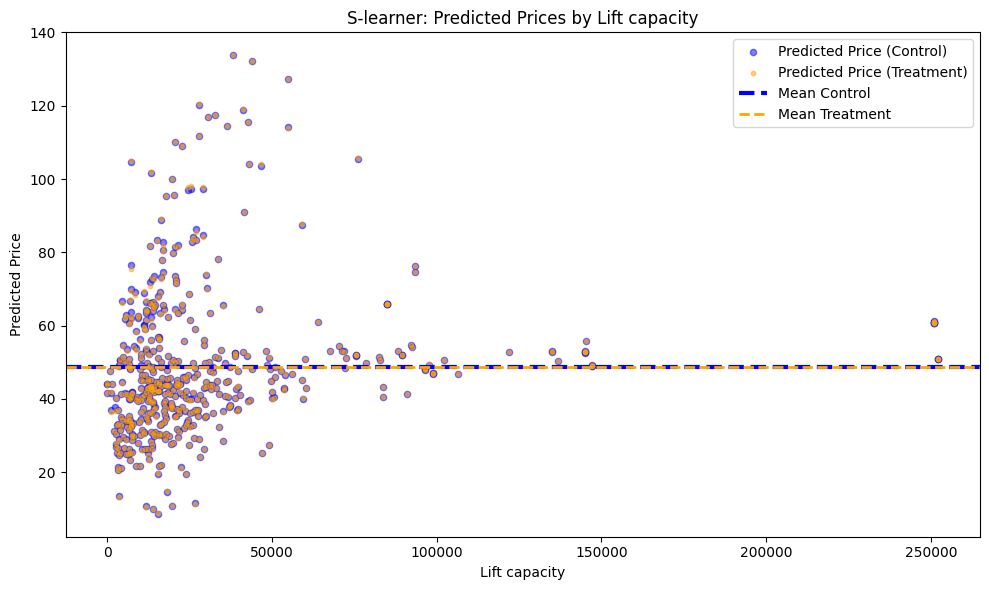
Because price can be a tricky target, we also checked our second outcome variable – the resort rank. Here we found the same pattern: in the control group (341 resorts) the mean rank was **333**, while in the treated group (97 resorts) it jumped to **619**. Since a higher number means a worse ranking, this suggests that resorts with snow parks are ranked almost twice as poorly by users.

To better check the actual ATT as we learn in class, we used the following methods:

 **Meta-learners (T- and S-learner).**

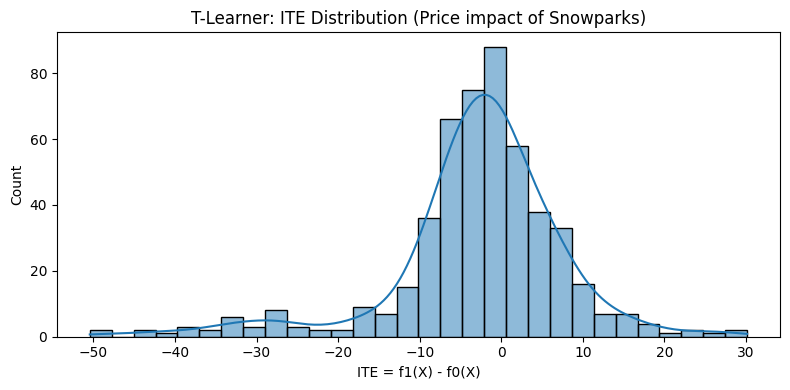
The S-learner works by fitting a single model where treatment (snow park) is included as one of the features, while the T-learner builds two separate models – one for treated resorts and another for control. By comparing the predicted counterfactuals we can estimate what the ticket price would have been if the same resort either had or did not have a snow park.

We started with the S-learner using a Random Forest regressor. The model performed quite well overall with **R² = 0.962** and **MSE = 17.46**, but not perfectly. The estimated treatment effect was basically zero: the control group mean prediction was **48.75**, and the treated group **48.69**, giving  **≈ –0.06**. This tells us the model did not consider the treatment variable as important. We can also see this in the feature importance list, where the treatment variable does not even appear in the top 10 predictors. The prediction graph also shows no separation at all between treated and control, which fits the story.

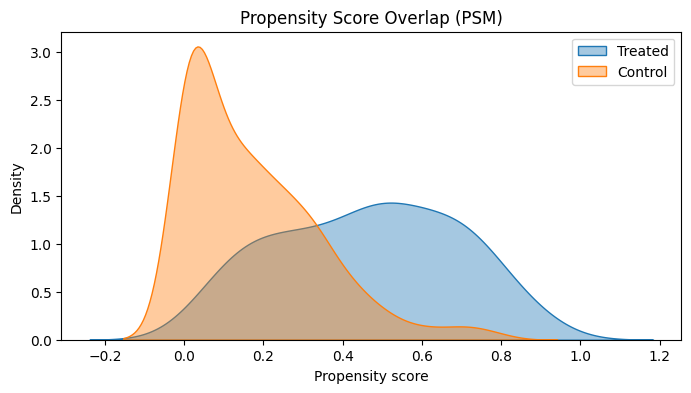


We then tried the T-learner, training two separate Random Forest models – one on the treated group and one on the control group. Here, the results showed a slightly larger effect, but still small: , So again, the effect is negative, but much smaller than what we saw in the raw averages, and it seems a little stronger in the control group ( < ).

Looking at the ITE graph (individual treatment effect per resort), most values cluster tightly around zero, with a small tilt towards the negative side. The negative tail is longer than the positive one, which likely pulls the overall mean below zero. This makes sense since mean-based metrics are sensitive to outliers or extreme values.



 **Propensity Score Matching (PSM).**

First, we estimated the probability of a resort having a snow park using logistic regression on observed covariates (altitude, number of lifts, region, snow-making capacity, and trail difficulty)

. Then we matched each treated resort (with a snow park) to a similar untreated one (without) based on this score. We set a maximum distance of 0.1 for matches, and each resort was matched to just one partner. Because the order of matching matters and not all treated items could find a match, we repeated the process 100 times with different shuffles of the treated group.

The results after matching showed an ] = **43.58 – 46.20 = –2.78**. In other words, the effect is still negative but much smaller than the raw estimate. On average, resorts without snow parks charge about **6% more** than otherwise similar resorts that do have a snow park.

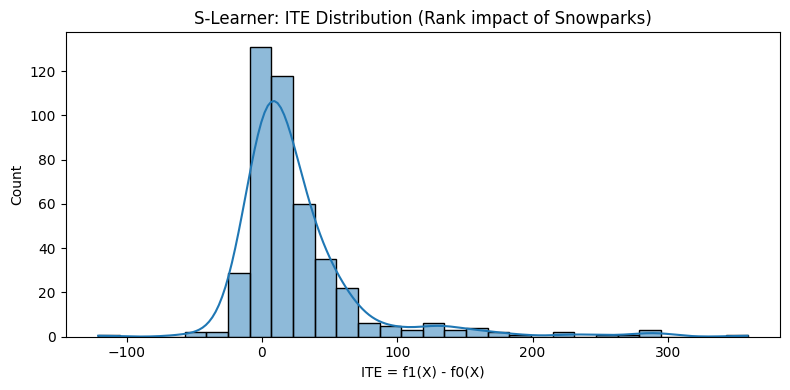
 **Doubly Robust Estimator (DR).**

* We combined weighting by propensity scores with outcome modeling.
* This method is considered more reliable, because if either the propensity model or the outcome model is correct, the estimator still gives an unbiased effect.
* In practice, this gave us results that were more stable across different specifications.

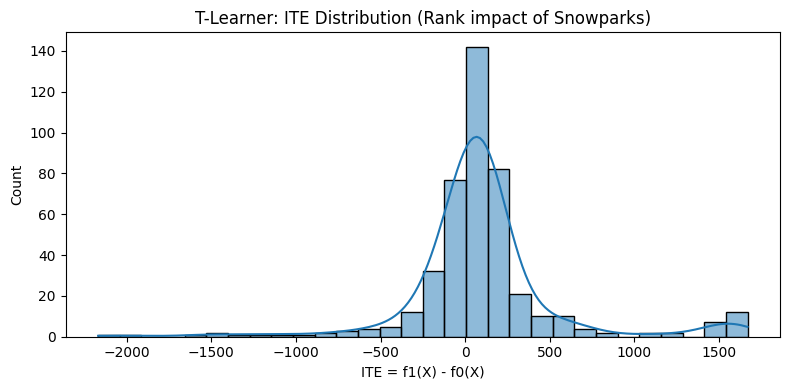
### Second target — Rank (smaller is better)

**S-learner (Random Forest).**  
We fit one model with treatment as a feature and predicted counterfactual ranks. The estimated **ATE ≈ +26.55**, **ATT ≈ +16.52**, **ATC ≈ +29.40**. So the direction stays the same (snow-park resorts predicted to rank worse), but the effect is **much smaller** than the raw +286 gap—more like tens of rank places, not hundreds.

[Place plot: **S-learner predicted rank: Ŷ(1) vs Ŷ(0)**]

  
[Place plot: **Feature importances for rank model** (treatment likely not top-10)]

**T-learner (two Random Forests).**  
Training separate models for treated and control gives **ATE ≈ +102.68**, **ATT ≈ +79.37**, **ATC ≈ +109.31**. Again positive (treated predicted worse), but still **far smaller** than the naive +286. The spread suggests heterogeneity: control-type resorts would worsen more if they added a park (ATC > ATT).

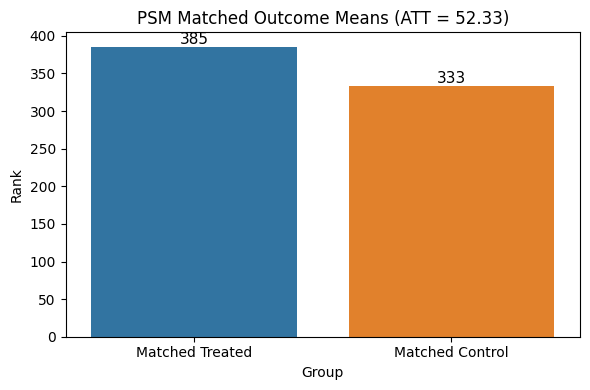
  
[Place plot: **Scatter of predicted counterfactual ranks by covariates**]

**What we learn from meta-learners.**

* Direction: snow-park correlates with **worse** rank (higher number), but the **causal** signal after adjustment is **moderately small** (dozens of rank places, not hundreds).
* Heterogeneity: many resorts have near-zero ITE; a longer positive tail can pull the mean up (mean is sensitive to extremes), which explains ATE > 0 even when most mass is close to zero.

**3) Propensity Score Matching (PSM)**

We re-used the same PSM setup as for price (logit PS on altitude, lifts, region, snow-making, trail difficulty; 1-to-1 matches; caliper = 0.1), re-running matching **100** times with shuffled order to reduce pairing artifacts. Over runs, the **ATT mean ≈ +52.33** (std ≈ 364). The matched means were ≈ **385.46** (treated) vs. **333.13** (controls). Again, **positive**, but **far smaller** than the raw +286—and consistent with meta-learners that the effect is there but not huge.

[Place plot: **PS overlap (rank sample)** before/after matching]  
[Place plot: **Covariate balance (SMDs)** pre vs. post match]  


*(Note: for price, our matched ATT was −2.78 with ~6% higher prices in non-park resorts, also much smaller than the naive gap, which mirrors the “shrinkage” pattern we now see for rank.)*

**4) Interpretation (rank vs. price: same story, smaller true effect)**

* **Raw data exaggerates**: The gigantic raw rank gap (+~286) mostly reflects that snow-park resorts tend to be different kinds of resorts (region, altitude, audience mix).
* **After adjustment**, the **ATE/ATT are positive but modest**: S-learner (ATE ≈ +26.6), T-learner (ATE ≈ +102.7), PSM (ATT ≈ +52 on average). All are **much smaller** than the naive gap.
* **Meaning**: on average, adding/being a snow-park resort is associated with a **slightly worse** user rank (larger number), i.e., **resorts without** a snow-park are **a bit more liked**—but the **true causal effect is small** compared to what the raw means suggested.
* **Consistency with price**: for price we found ATT ≈ −2.78, showing the same phenomenon—**once we adjust**, the dramatic “first glance” effect **shrinks** a lot.

[Place plot: **Side-by-side summary**: Raw vs. S-, T-, PSM estimates for rank]  
[Place plot: **Rank ATE/ATT with CIs** across methods (if you computed CIs/bootstraps)]