Microeconomics with Ethics

by Steven Suranovic (2022) – George Washington University

Chapter 10
Monopoly Markets

In the next few chapters we will consider several different market structures on the supply side. We begin here with the structure of a *monopoly firm*, defined as one firm supplying a product in a market with many buyers, facing no competition from other firms selling identical products or close substitutes. The monopoly market is the easiest to analyze and is a good starting point to address the effects of competition as we progress through the increasingly complex market structures of duopoly, oligopoly and perfect competition in subsequent chapters.

10.1 What are Monopolies?

Learning Objectives

1. Learn the definition of a monopoly and the three basic reasons they may arise.

A *pure monopoly* occurs when there is only one firm supplying a product to a market and there are no close substitutes that consumer may purchase instead. The second feature makes it very difficult for any firm to establish a pure monopoly. One classic example of a monopoly is electricity generation. Most people can buy power from only one electric company that operates in a local region. However, there are some close substitutes available. A household can purchase a generator which uses a gasoline powered engine to produce electricity, or, they could install solar panels to produce electricity. That means the electric company is not a pure monopoly, but is sometimes called a *near-monopoly*.

Economists sometimes also say a firm has *monopoly power, or market power,* to describe a company that is close to being a monopoly, such as with electric companies, but is not a pure monopoly. To say that a firm is trying to acquire monopoly or market power means they are attempting to eliminate some competition in order to be able to operate more like a pure monopoly. We will discuss many of the methods firms use to acquire more monopoly power in Chapter 16.

Three Reasons Monopolies Arise

There are three distinct reasons, monopolies, or near monopolies, can occur; 1) because inputs are monopolized, 2) because the product has features associated with a natural monopoly, and 3) because the government grants monopoly privileges, or rights.

Case 1: Monopolized Inputs

Perhaps, the best example of monopolized inputs creating monopoly power is the case of the De Beers diamond company, founded by Cecil Rhodes and Barney Barnato in South Africa in 1888 and named for two Dutch brothers who originally owned their first diamond mine. Their strategy was to buy and control as many of the diamond mines as possible around the world, so if you needed a diamond, you had to go to De Beers. According to Wikipedia, De Beers controlled over 80% of the diamond market for over a century between its founding until the start of the 21st century. By 2020, competition had reduced its toehold and its market share had plummeted to less than 30%.

The De Beers company is also responsible for promoting diamonds as a symbol of love and the only acceptable precious stone to be used on an engagement ring. This advertising campaign began in the 1930s and resulted in the popular 1940s advertising slogan, Diamonds are Forever.

A second example of a monopoly arising from control of the inputs in production is the case of the Ambassador Bridge connecting Windsor, Ontario, Canada with Detroit, Michigan in the US. Unlike most roads and bridges in the US and Canada, the Ambassador bridge is privately owned. The product being produced with a bridge is not a good, but rather, a transportation service for cars and trucks across the river and the expenses to build the bridge and for maintenance are paid for from the tolls collected for passage. The Ambassador bridge opened in November 1929 and only had a pure monopoly for one year when the Detroit-Windsor tunnel, "co-owned" by the city governments of Detroit and Windsor, and a very close substitute to the bridge service, opened in November 1930. Thus, the bridge has been, at-best, a near-monopoly for close to a hundred years.

Case 2: Natural Monopolies

A second reason monopolies sometimes arise is because the product or service being produced has a characteristic that makes it a natural monopoly. A *natural monopoly* occurs when production exhibits economies of scale up to a level of production that is adequate to supply the entire demand in the market. Recall that it is expected that most production processes will achieve minimum average cost at some "intermediate" level of production. For most products that intermediate level will be much smaller than the total market demand for the product which will enable many separate firms to enter the market and compete with each other. However, for some products, especially ones whose production has a very high level of fixed costs, the minimum average cost might occur at a level much larger than the total market demand. This would mean that if a firm can grow larger than a competing firm, it will be able to lower its costs relative to the other, lower the price it charges, and force the competitor out of business. Thus there is a natural tendency for a monopoly to form.

The best example of a natural monopoly is electricity generation and supply. The source of the high fixed costs are the copper wires that must be extended to every business and household connecting them directly to the electric power generating plant. These are fixed costs because this expense must be incurred and financed before even one watt of electrical power is delivered. Also, if power generation were to drop to zero one week, the payments to service the loans that financed this electrical grid would still have to be paid. As a result, one company suppling power to a region can do so at a much lower unit-cost than if multiple firms tried to compete for customers. As we will see later, the inclination for a monopoly firm to set very high prices for their products has led to government regulations that control the price that electric power

companies can charge their customers. These regulations, in effect, prevent a natural monopoly from behaving like a monopoly.

A few other examples of products with high fixed costs and thus conducive to being natural monopolies include other utilities such as wired telephone, natural gas, water, and cable television.

Case 3: Government-Granted Monopolies

Finally, the third way a firm might obtain a monopoly is because the government grants that privilege to it. There are two subcategories in this situation though: 1) the government might take control of an industry and create a state-owned enterprise, and 2) the government can grant intellectual property rights in the form of a patent and give exclusive rights to the patent holder to produce and sell that product.

State-owned enterprises (SOEs) are common in countries that have authoritarian governments because by controlling and monopolizing an industry, the government can earn monopoly profits (discussed later in this chapter) and directly benefit the government officials themselves. However, not all state monopolies are operated to maximize profits. Some SOEs are run like non-profit institutions with a goal to satisfy the customers as best as possible. These type of SOEs face the problem that if profits are negative, they cannot go out of business and so the losses must be covered by tax revenues collected from the citizens.

As examples, China has many SOEs because in their transition out of central planning, which prevailed from 1949 until about 1978. Although they allowed for the operation of private businesses, they maintained control over some firms in some industries. Many of these SOEs still exist today, some 40 years later.

Another example is drawn from the history of the UK. After the Second World War, Britain nationalized a number of different companies in the coal, steel, and transportation sectors. The objectives of nationalized firms often are not to maximize profit but rather to pursue other social objectives, an important one being, the conditions of workers including adequate wages and high levels of employment. Sometimes these SOEs actually compete with private firms in the same industry and thus face some degree of competition.

The second way that government creates monopolies is through the provision of intellectual property rights (IPR) in the granting of patents. A patent can be awarded to an innovative product that is uniquely different. The patent provides the holder with the exclusive right to produce the product for a period of 20 years. In other words, it grants a monopoly. The patent holder could operate as a monopoly for those 20 years, and earn monopoly profits, or they can license the right to produce the product to others and obtain a *royalty payment* in exchange. If someone who is not the patent holder produces the product themselves, the patent holder can sue them using the government judicial system and force them to cease production or pay compensation.

The reason IPR is granted by governments is largely to promote innovation. If someone knows that a new invention could give them a guaranteed high return in the future, then they are more likely to seek to innovate. Monopoly privileges are viewed as a necessary allowance to businesses in order to promote a better outcome in the long term.

Key Takeaways

1. There are three main reasons monopolies may form
    a. because production inputs are monopolized,
    b. because the product has features associated with a natural monopoly
    c. because the government grants monopoly privileges, or rights.

10.2 Monopoly Market Analysis

Learning Objectives

1. Learn how a monopoly would choose the quantity to produce and the price to charge to maximize its profits.
2. Identify the profit-maximizing condition for a monopoly firm.
3. Learn to derive and interpret the marginal revenue curve.
4. Depict total revenues, total costs, and total profit for a monopoly firm in a demand-cost diagram.
5. Learn how to measure market welfare in a monopoly market.

In this section, we evaluate how a monopoly firm, operating in a market with many potential consumers, would decide how much of the product to produce and what price to charge. We will assume that the monopoly is a private firm, with one or more owners, whose objective is to maximize profit in operating this business. We will assume that the monopoly firm has a production function that exhibits economies of scale at low levels of output, diseconomies of scale at higher levels of output and has some level of fixed costs in production. This implies that their average and marginal cost functions will have the patterns like those presented in Figure 9.2. We will not name the product in this exercise, but will assume it can be measured in pounds and is priced in $/lb.

Profit Maximization

We will assume the monopoly firm owners have perfect information about the market demand function. This means they know precisely how much of their product they will be able to sell at every possible price they might charge.

As introduced in the Chapter 9, profit can be written as the following equation:

$$\Pi = TR(Q) - TC(Q)$$

where TR(Q) is total revenue written as a function of the quantity produced, Q, and TC(Q) is total costs also written as a function of Q. We can extend this equation by plugging in the market demand function. For example, suppose the market demand function can be written as $Q_D = F(P)$ with Q_D representing the quantity demanded and F(P) being some function that includes the price P as a variable. If we knew the precise functional form we could rewrite the equation with P on the left-side and Q on the right-side. In general, we can do this by writing , $P = F^{-1}(Q)$, where $F^{-1}(.)$ is the inverse demand function. Finally since TR = PxQ we can write the total revenue function with the demand curve included as $TR = F^{-1}(Q) \times Q$ and the profit function becomes,

$$\Pi = F^{-1}(Q) \times Q - TC(Q)$$

which explicitly includes the information from the market demand function and the information about production costs in the profit equation for the firm.

What economists do next is to use a little calculus, to determine the quantity that would maximize the value of this function. We won't do that here, but what we will do is to show the condition that comes from that mathematical exercise.

In particular, the math tells us that to maximize profit the monopoly firms should choose a quantity of production, Q, such that,

$$MR(Q) = MC(Q)$$

Where MR is the marginal revenue and MC is the marginal cost, both written as a function of Q. Since the optimal Q plugged will be the same in both functions we will often simplify this and say, find Q such that MR = MC.

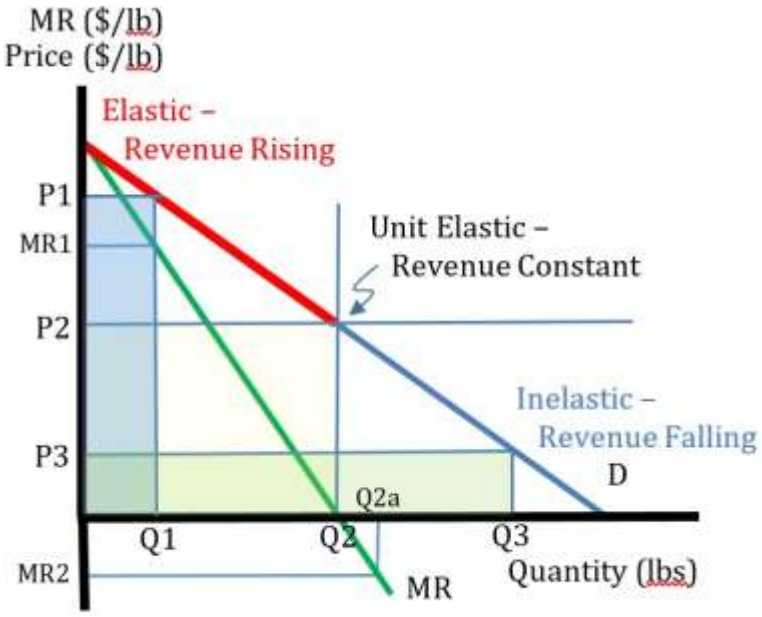
The marginal cost function for a typical firm was derived in Chapter 9 and graphed in Figure 9.2. Recall that it is U-shaped but upward sloping as it passes through the minimum average cost function. Stay tuned, the relationship between marginal cost and average cost will be important in this and subsequent chapters.

The Marginal Revenue Curve

Marginal Revenue is defined as the change in total revenue given a change in the quantity supplied. It can be written as $MR = \Delta TR/\Delta Q$. Note that since a change in revenue is in the numerator and is measured in $, and Q is in the denominator and is measured in units such as kilograms, then marginal revenue is measured in, for example $/kg. This is good because marginal cost is also measured in units like $/kg and if these two variables must be equal to each other, they have to be measured in the same units.

We introduced total revenue in Chapter 8 and depicted in Figure 8.2 how revenue could be measured using a market demand curve, how total revenue varied along the demand curve, and how revenue changes are related to demand elasticities. We recreate Figure 8.2 below and add information about the marginal revenue curve in Figure 10.1

Figure 10.1 The Marginal Revenue Curve



The marginal revenue curve is the green line in Figure 10.1 and can be easily derived and drawn when market demand is linear. In this case, the MR curve (or line) will intersect the demand curve on the vertical price axis, and will intersect the horizontal axis at the quantity associated with the demand midpoint. We'll explain the reasons in a moment.

But first, note that marginal revenue is measured along the vertical axis since it has units of $/lb in this case, and those are the same units in which prices are measured. MR is a function of quantity, measured along the horizontal axis. Note that this is the reverse of the demand curve. Demand is a quantity measured along the horizontal axis in pounds and is a function of the price which is measured along the vertical axis. This is sure to cause some confusion, if not recognized.

As discussed in Chapter 8, total revenue rises as the price falls from P1 to P2 where the demand elasticity is elastic. This means that marginal revenue must also have a positive value in that range. At quantity Q1, any increase in quantity, and decrease in price, along the demand curve, will cause revenue to rise. Hence marginal revenue at Q1 is in the positive range at the value MR1 listed on the vertical axis. When the quantity is at the midpoint of the demand curve, Q2, total revenue is maximized and demand is unit-elastic. Any increase in quantity above Q2 would therefore result in a lower total revenue and thus marginal revenue would be negative in value. This is why at quantity Q2a, marginal revenue is negative at the value MR2. If the quantity is changed from just a little below the revenue maximum at Q2 to a little above Q2, then revenue would not change. That is why marginal revenue is measured as zero at Q2. Finally, and for completeness, let's imagine that the first unit of output on the diagram is an extremely small distance above zero along the quantity axis, perhaps the distance of a human hair width. To sell that first unit, a firm could charge a price essentially equal to the price where the demand curve intersects the price axis. Let's call that price, P_H. (not shown on the diagram). Total revenue earned on that first pound of output would be $P_H \times 1 = P_H$. Hence the MR at the

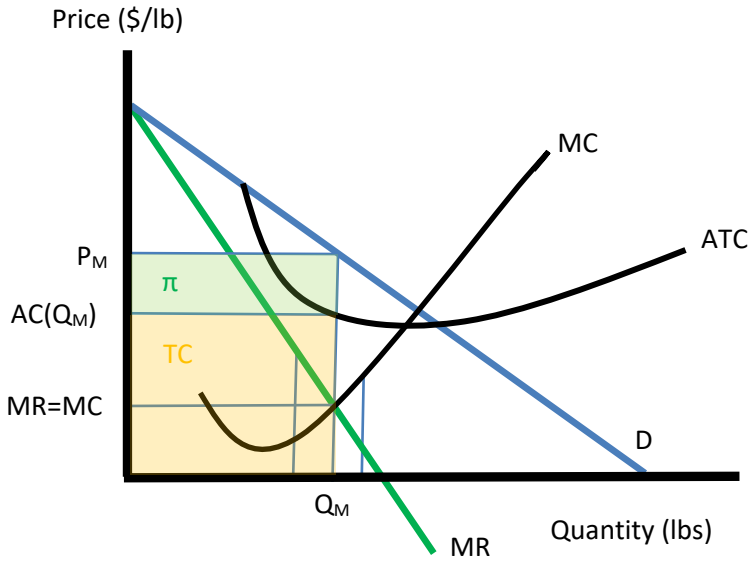
quantity zero, must be equal to P_H. This is why MR intersects the demand curve on the vertical axis.

To summarize, for all quantities between zero and the midpoint Q2, (equivalently, for all quantities in which demand elasticity is positive, or elastic), marginal revenue is also positive. That means that any increase in quantity in this range (or decrease in price), will cause total revenue to increase. Equivalently, any decrease in quantity (or increase in price) in this range, will cause total revenue to decrease. Marginal revenue falling in this range (the downward slope of the MR curve) means that the rate at which total revenue increases with rising quantity, although positive, is falling. Stated differently, a one pound increase in quantity near Q1 will raise total revenue by more than a one pound increase in quantity near Q2.

Illustrating the Monopoly Decision

Next we can merge the monopoly firm cost functions with the market demand curve in Figure and its associated marginal revenue curve in Figure 9.2 to illustrate the monopoly firm's profit maximizing choice of output level and price.

Figure 10.2 Monopoly Profit Maximizing Price and Quantity



Notice that the marginal revenue curve, MR, intersects the marginal cost curve, MC, at the quantity Q_M. This means Q_M is the profit-maximizing output level for the monopoly firm.

To explain why, suppose the firm instead chose to produce at a level less than Q_M, such as where the vertical line is drawn (but not labeled) in the Figure. At that output, MR (measured along the vertical axis) > MC (also measured vertically) which, applying the definitions, means that a one unit increase in output would raise total revenue by more than it raises total cost, and since profit = TR – TC, profit would rise. Similarly, suppose the firm instead chose to produce at a level greater than Q_M, such as where the vertical line is drawn (but not labeled) in the Figure. At that output, MC (measured along the vertical axis) > MR (also measured vertically) which means that a one unit increase in output would raise total cost by more than it raises total revenue, and since profit = TR – TC, profit would fall. Thus, to maximize profit, the monopoly firm would choose to produce Q_M.

The price the monopolist sets is read off of the demand curve. In order to sell Q_M of the product, the firm should set the price at P_M. Thus, P_M is the profit maximizing price for the monopoly firm.

Total Revenue, TR = $P_M Q_M$, and is depicted in the Figure as the entire shaded area which is the sum of the green and yellow areas. The yellow area represents Total Cost at output level Q_M. This is found by finding where the quantity line at Q_M intersects the average total cost curve, ATC, to find average cost at the quantity Q_M, $AC(Q_M)$, which is read on the vertical axis. Because AC = TC/Q it implies that AC x Q = (TC/Q)Q = TC. This is what the yellow shaded area in the Figure is Total Cost when quantity Q_M is produced. Finally, since the whole shaded area is TR, and the yellow area is TC, profit, or π = TR – TC, and is depicted as the green shaded area in the Figure.

SIDE NOTE: Notice that two other ways to write the profit relationship is

$$\pi = P(Q) \times Q - AC(Q) \times Q \quad \text{or}$$

$$\pi = [P(Q) - AC(Q)] \times Q$$

which will work for any Q that is plugged in, including Q_M. To find profit for a particular quantity, draw a vertical line at the quantity up to the demand curve to find P(Q) and use that same line to find its intersection with ATC to find AC(Q). The difference between these two values, measured on the vertical axis, times the quantity Q, yields firm profit at quantity Q.

Market Welfare

Market welfare is measured as the sum of surplus accruing to the consumers who purchase the product plus the surplus that goes to the firm in the form of profit. The monopoly profit has already been determined as the green area in the Figure. Recall from Chapter 8 that consumer surplus is measured as the triangular area bounded by the demand curve, the price line at P_M, and the vertical axis. Thus, market welfare is the sum of the green profit rectangle and the consumer surplus triangle.

In this monopoly market, Q_M of the product is bought voluntarily at the posted price of P_M by consumers who are willing to pay at least P^M, or more, to acquire the product. Q_M is also sold voluntarily by the monopoly firm generating surplus profit given by the distance (P_M – $AC(Q_M)$) for each unit sold. Both consumers and the owners of the monopoly return home happier than before they went to the market.

Key Takeaways

1. A monopoly firm would maximize its profit by choosing the output quantity that equalizes marginal revenue with marginal cost.
2. The monopoly price is the price on the market demand curve associated with the profit maximizing quantity.
3. The profit maximizing condition is MR = MC
4. Marginal Revenue is the change in total revenue given a change in quantity.
5. The marginal revenue curve for a linear demand curve is also linear, intersects the demand curve at the vertical axis and intersects the horizontal axis directly below the midpoint of the demand curve.
6. Market welfare in a monopoly market is the sum of firm profit plus the consumer surplus arising at the profit maximizing price and quantity.

10.3 Price Discrimination

Learning Objectives

1. Learn how a firm can use price discrimination to shift surplus from consumers and raise its profit
2. Identify real-world examples of price discrimination

In this section, we demonstrate that a monopolist could actually make even more profit than the profit-maximizing level derived above in Section 10.2. How could that be possible? How can one obtain more than the "maximum?" Well, the answer comes by relaxing, i.e., changing, one of the behavioral assumptions in the previous model. In the previous model we assumed, unquestioningly, that the monopoly firm would set one price and charge every consumer that same price. But what if instead the monopoly firm could arrange, somehow, to change different consumers different prices. Any pricing scheme that charges different consumers different prices for the same product is called *price discrimination* and it occurs in the world much more often than you might at first think.

But before offering examples of real world price discrimination, let's first examine why it can raise a firm's profit above the previously identified maximum. The best way to do that is to first imagine the case of what economists call perfect price discrimination. In *perfect price discrimination,* the firms charges each and every consumer a unique price, a price tailored to the characteristics of that consumer. It is virtually impossible to achieve this goal, but here is how it could play out, hypothetically.

Suppose the monopoly is at first operating to maximize profit with, producing quantity Q_M and setting price P_M (let's say at $4 per pound) as derived in section 10.2. The firm knows that it will sell all of its product if it charges the same $4/lb for everyone. However, imagine further, and outlandishly, that the firm has developed a truth serum that will elicit a truthful response to any one question asked of the consumer when checking out to purchase their product. Having administered the serum, the seller could ask, "how much are you willing to pay for the product?"

Now recall that the demand curve in Figure 10.2 shows that all of the consumers who purchase the product at $4 would actually be willing to pay more. That extra amount is the surplus value the consumer receives had she purchased it at the price $4. Thus each consumer who is asked the truth-induced question, how much are you willing to pay?, will respond with an answer unique to them. One person might say $5 and the merchant could than charge him $5. Another might say $4.50 and seller could sell it to her at that price.

If the seller could implement such a scheme, then the monopoly firm would be able to capture all of the consumer surplus as measured by the triangular area between P_M and the demand curve in Figure 10.2. And this is on top of the profit already indicated in the previous exercise. Hence by implementing perfect price discrimination the monopoly firm could increase their profitability.

Of course there are many problems with this model that make it unworkable in real life. First there is no truth serum. Second, you would have to offer a slightly lower price than each person's willingness-to-pay because otherwise they would be indifferent between purchasing it or not. And third, you would have to hope that consumers would accept such obvious discrimination. In addition, even if one were able to perfectly price discriminate, the monopoly could do even better by selling a greater quantity than Q_M at even lower prices for those

consumers only willing to pay less than P_M. Indeed, the optimal solution is to produce the quantity where MC crosses the demand curve and charge every consumer their individual willingness-to-pay price.

Practical Price Discrimination

The point of the previous unrealistic model was to highlight to a potential for a firm to grab onto a larger share of the surplus that accrues to both parties in an exchange. With perfect price discrimination a firm could greatly add to their own profit surplus by taking away all of the surplus from consumers on the other side of each trade. Since we imagine that firms seek to make as much profit as possible, we should also recognize that if a firm could achieve this outcome, they would.

Nonetheless, firms have come up with some creative alternatives that enable them to capture some of the consumer surplus and thereby raise their profitability. One such method separates potential consumers into distinct groups by recognizing that the demand curve for each group varies with respect to their elasticities of demand. For example, the demand for movie theater tickets by more elderly consumers and by families with children is likely to be much more elastic than demand by the average adult. That's because these consumers tend to be more cash constrained and price sensitive and so their willingness-to-pay will likely be lower than for the average adult. Movie theater companies have responded by offering lower priced tickets for individuals over 60 or under 12. This is price discrimination because different customers of the same service are charged different prices. It also enables the companies to raise their profits because the lower prices offered to the price sensitive consumers will increase their demand enough to offset the lower demand by the price insensitive consumers who now pay higher prices. And yes, adults pay higher prices than they would otherwise if the firms had not engaged in price discrimination.

Another classic example is in the pricing of airline tickets. This practice may have lapsed, but at one time it was generally cheaper to buy a roundtrip ticket from say Chicago to New York, if the flight departed on Friday and returned on Sunday or Monday, than for a flight that departed on Monday and returned on Thursday or Friday. The logic was that "Saturday stayover" flights were more likely to be purchased by consumers visiting relatives who were thereby more likely to be price sensitive, (elastic demand). In contrast, weekday flights were more likely to be business travelers whose companies were less likely to be price sensitive (inelastic demand). It's the same outcome as with movie tickets. The higher demand by weekend travelers was enough to offset the slightly lower demand by business travelers who also paid higher prices than otherwise. What's especially ingenious about this arrangement is that it never required asking a customer whether they were traveling for business or pleasure.

There are many other examples of price discrimination including offering , and much more. There are no explicit laws against price discrimination and we make no arguments here about the overall goodness or badness of the phenomenon. Instead, we merely intend to illustrate the practice and demonstrate that it represents methods to shift some of the surplus that arises in trade away from the consumer and towards the producer.

Key Takeaways

1. Price discrimination occurs when a firm charges different prices to different customers.
2. In perfect price discrimination a firm can shift all of the consumer surplus to increase their profit.

3. Some examples of real-world price discrimination include discount movie tickets, Saturday-stayover airline ticket discounts, discount coupons, college tuition discounts, bulk purchase discounts, premium pricing, military discounts, ladies' nights in clubs, and academic pricing for software.